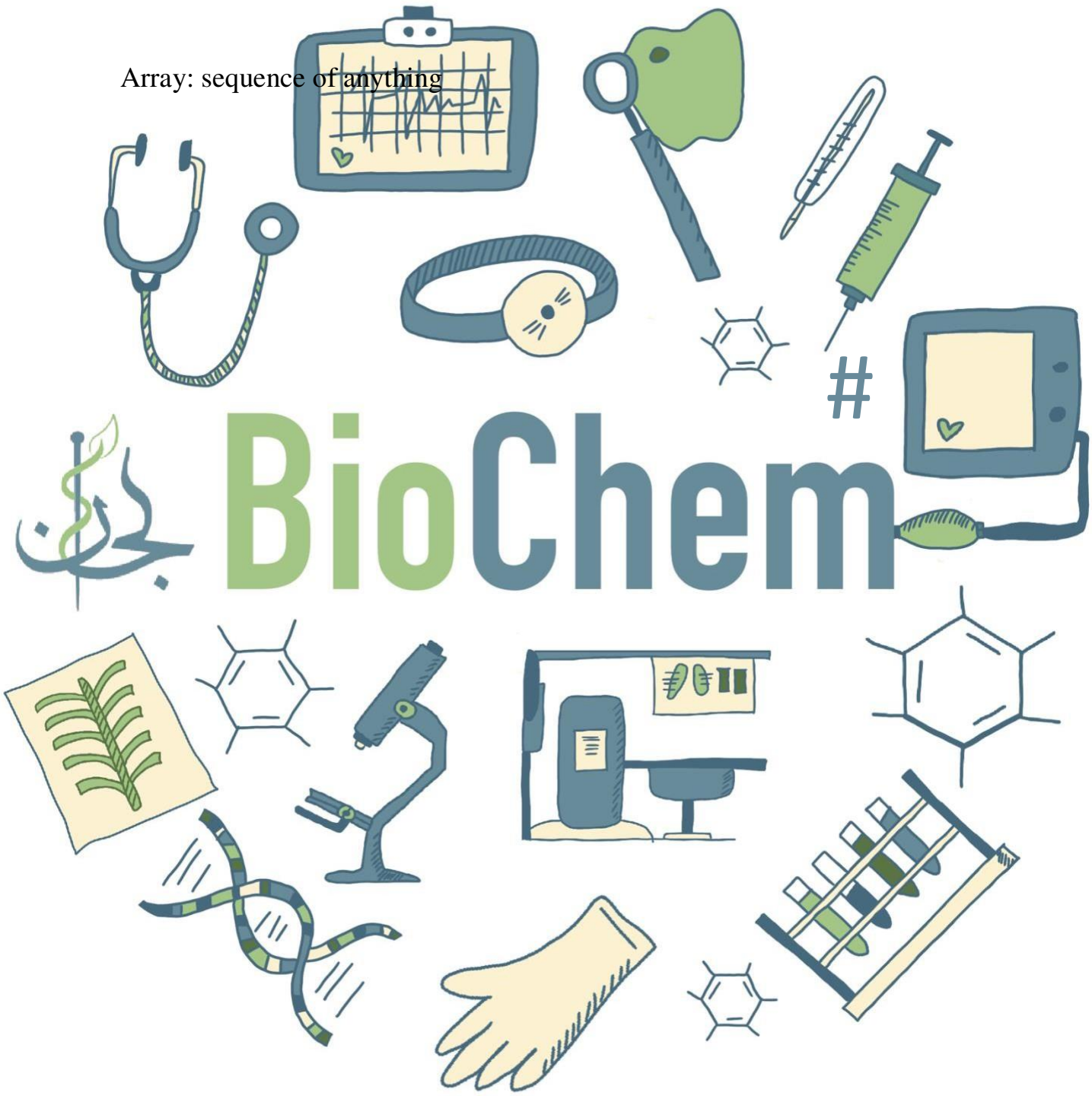
Array: sequence of anything



Writer:	Abdelhadi Okasheh / Yamama Alabbadi
Science:	Hani Shihadeh
Grammar:	Moath Fateh
Doctor:	Mamoun Ahram

ملاحظة صغيرة : الكلام الموجود في السلايدات موجود في هذا الشيت باللون الأحمر و كلام الدكتور التوضيحي باللون الأسود.

Analysis of gene expression.

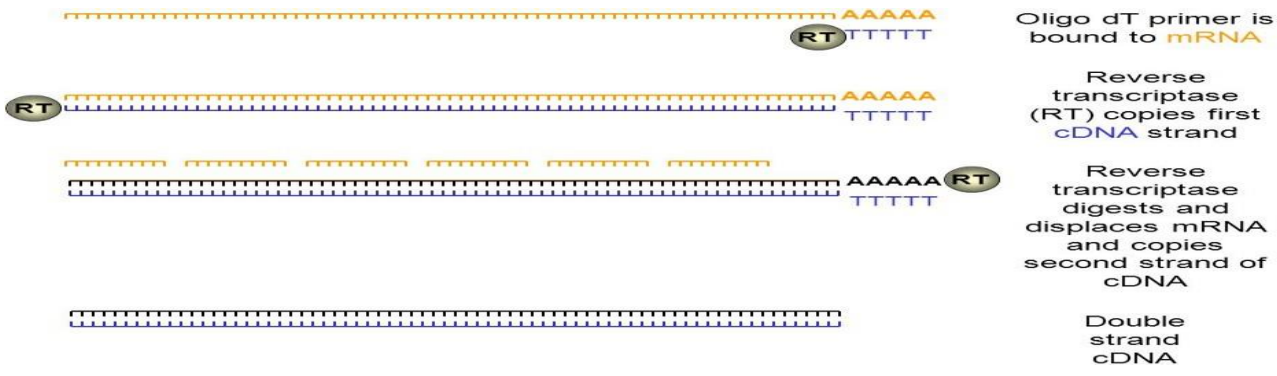
RNA level

- We are going to talk about Techniques leaded to analysis of gene expression via measuring the levels of RNA.
- in previous course we have learnt some basic techniques (northern blotting and in situ hybridization). With northern blotting we can know the relative expression of gene by looking at the thickening of RNA band and the size of mRNA.
- In situ hybridization we can also tell again the relative expression of a certain gene by staining the RNA molecules (you need tissue section) and you know where the expression takes place , in what cells of the tissue section .
- in this lecture we will cover three Techniques (two of them advanced and the last one is very advanced, the latest one).

❖ Basic methods: Northern blotting, in situ hybridization Advanced methods: real-time PCR, DNA microarray Very advanced methods: RNA-seq.

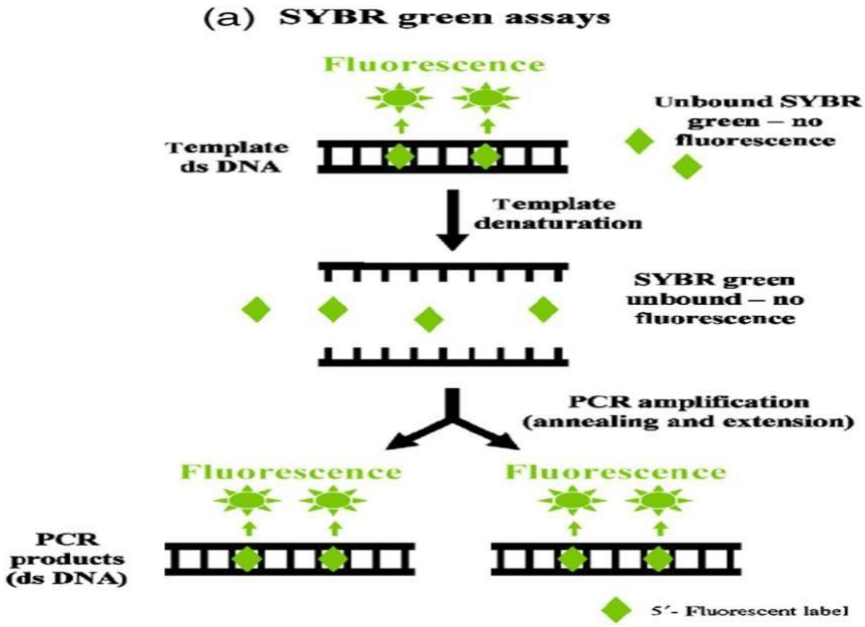
✓ It all starts by making a complementary DNA (cDNA) from mRNA using a primer **dT** (deoxy thymine) and this oligo(dT) will bind to the poly A tail of the mRNA ,the reverse transcriptase would make the first strand of DNA using mRNA as a template, and then the RT (reverse transcriptase) will digest the mRNA (the template strand) and synthesizes second strand, ending with a double stranded DNA.

✓ The relationship between the mRNA and the cDNA is (1:1) meaning that if we have 10 mRNA within the cell also we should have 10 cDNA molecules within the test tube.



the first technique : real-time PCR of cDNA

- Another way of relative quantitation of RNA expression is by converting RNA into cDNA followed by PCR in the presence of SYBR green.
- The higher the amount of RNA (cDNA), the sooner it is detected.

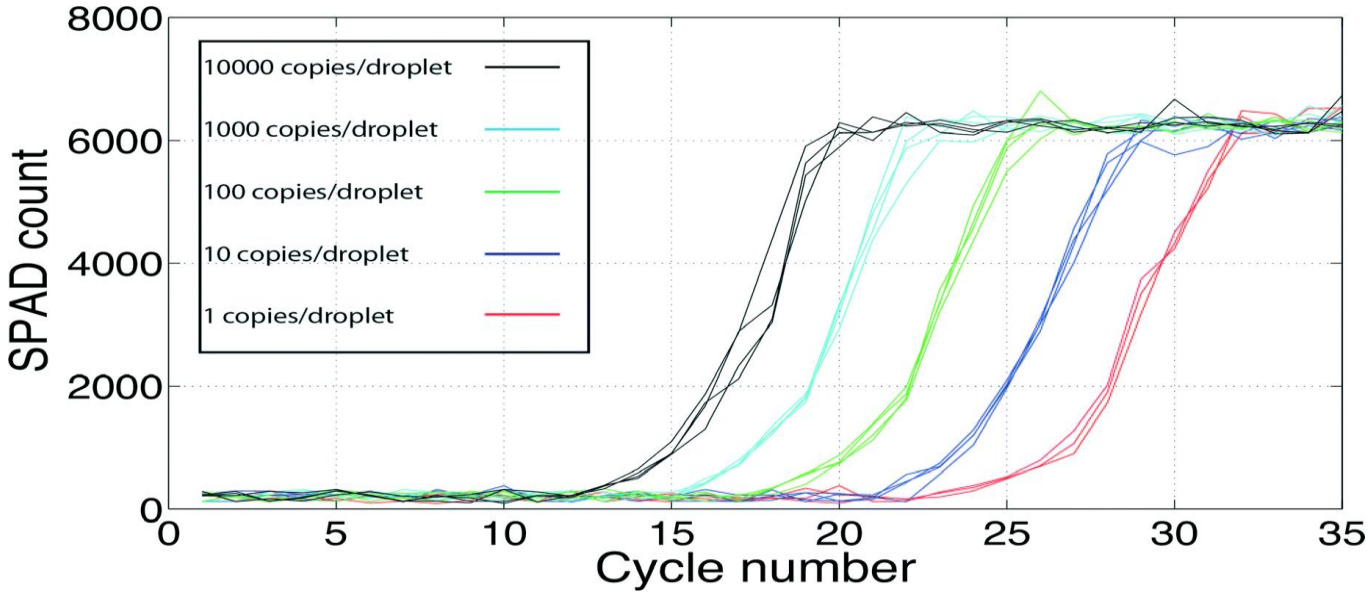


This technique could be quantitative meaning that we can quantify the number of molecules of cDNA/mRNA using real-time PCR we talked about real-time PCR before when we discussed a PCR and we talked about using SYBR GREEN (chemical that binds to double stranded DNA and then fluoreses)… the more the DNA in the sample the more florescence we will have.

What we do is:

1. Isolating mRNA molecules from cells
2. Converting RNA to cDNAs.
3. amplifying genes using specific primers.
4. We quantify (measure) the fluorescence generating after each cycle.

Remember: In early cycles there is nothing detected, THIS IS not because there is no amplification of cDNA actually there is an amplification (using a primer) that's because the instrument is not sensitive enough to detect very low amounts of molecules of cDNA.



- Here we have different samples range (1 copies-10000 copies) and the limit of detection id 100.000 copies (all samples would amplified for example sample one after the first cycle will be 2 then 4 … after the third cycle would be 8 and so on

  Sample 2: 10..20..40..80..160..etc

  Sample 3: 100...200..400..800..1600..etc)

  Until they reach the level of error (100,000 copies) , so here the sample with 10000 copies will be detected first because it will reach the limit of error (100,000) firstly (at cycle 12 , maybe 13 for example we will start having detection for that sample) … then we will detect sample 4, 3, 2 and finally the sample with only one copy eventually it would be detected later on (in the term of the number of cycles)

- The black sample having larger starting materials than the blue one … using real time PCR we can know relatively which sample have more mRNA molecules as a starting material than the other samples (depending on which sample we detect first)

Remember:

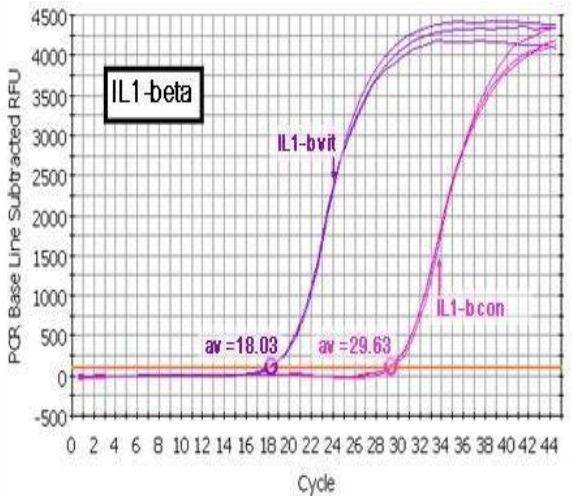
Housekeeping gene (gene that is needing to be expressed by the cell all time and its level is not different according to the environment

that cells live in like actin for example, tubulin…etc

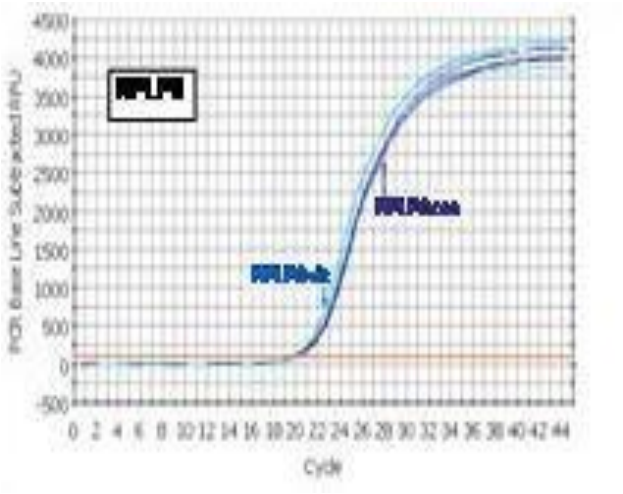
- We have to use the housekeeping gene to make sure that the amount of the starting material that I am using for different samples would be the same in terms of how much housekeeping gene there is, or in term of expression of the housekeeping gene (remember: that's the same idea of the northern blotting) …why I should do that?

  Because the expression of the housekeeping gene should not change, and that gives me a comfort in terms of the starting material that I'm using for samples that I want to compare.
- So when I try to do real-time PCR using two different samples to look at the relative expression of certain gene, I have to do analysis for the level of expression of the housekeeping gene like the genes producing actin, tubulin…etc to ensure that we are starting with a similar amount of the total RNA
- For example lets say that we are using for sample 1 a concentration of 1 microgram RNA for the starting material and for other sample we are using 10 microgram of RNA as a starting material so here that's a huge difference and we will see that the gene that we are analyzing for expression, the samples with 10 microgram of the total RNA molecules would appear earlier by real-time PCR. Is that an accurate result?

  Of course no it's artifact and it is not an accurate assessment because we are using different amount of RNA , but if we are using 1 microgram of RNA for all samples that would be more accurate (again we have to measure the expression of the housekeeping expressing to ensure that it is equal for all samples)



A gene of interest

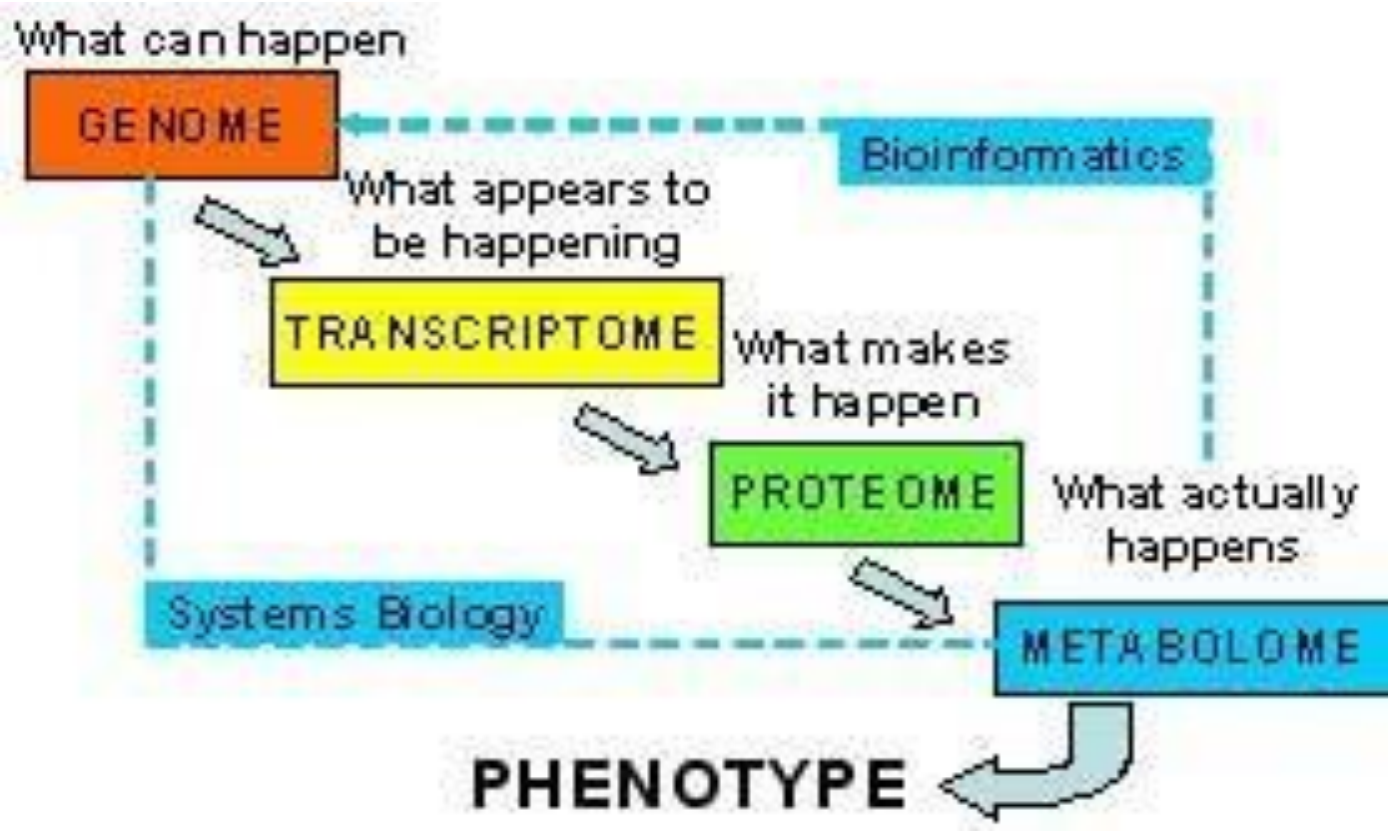


Housekeeping gene

- Look at the figure, by doing a real-time PCR of one housekeeping gene in this case a ribosomal proteins and its expression should not change according to condition, notice that for both samples (the dark and light blue) the detection of both samples shows same expression.

The science of -omics

- There are different technologies that started with an analysis of DNA and the analysis of the genome (the total collection of DNA in cell) and all the cells have the same genes and the same DNA and to look at what genes are important for what cells we have to look at the mRNA (its other name is transcript), so scientist who started to study mRNA (transcript) said: "since we have genome, we will call it (trascriptome) and the whole science of genomes is called (transcrpiotomics)".
- Other scientists who are protein biochemist said that what is really important for cells are proteins not mRNA – if we have 10 RNA molecules it doesn't mean that we will have 10 proteins – and the proteins not the RNA molecules do the function for the cells
- Then they come up with a new science known as (proteomics) which analyze the total collection of proteins within the cell.

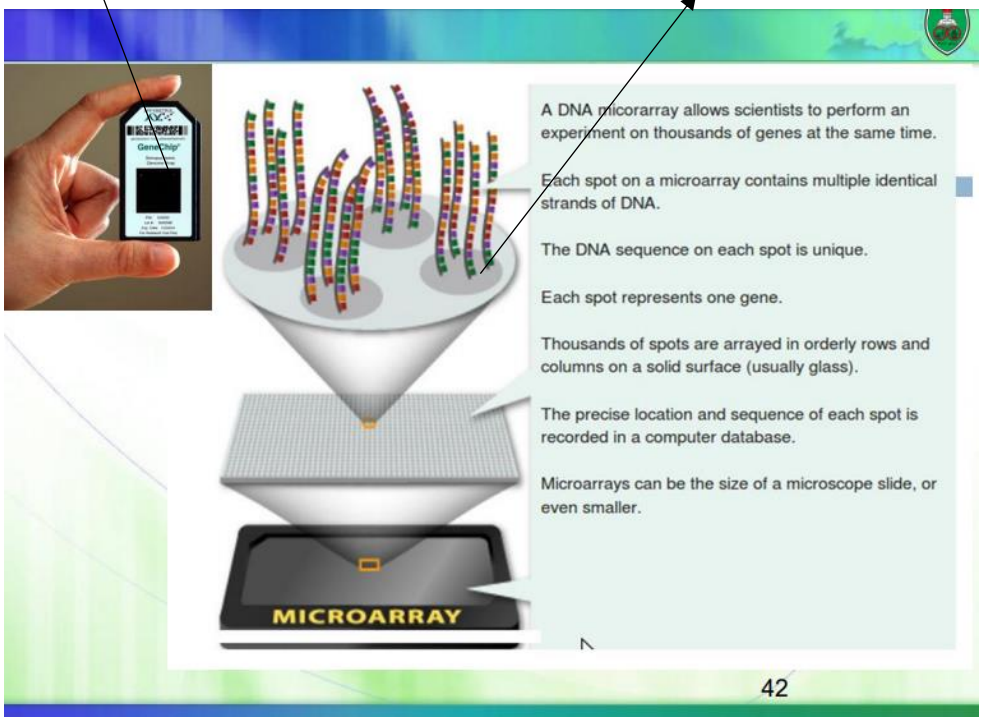


- One such method in studying transcriptomes is DNA microarrays, which allow the analysis of the RNA products of thousands of genes all at once.
- By examining the expression of so many genes simultaneously, we can understand gene expression patterns in physiological and pathological states.

.

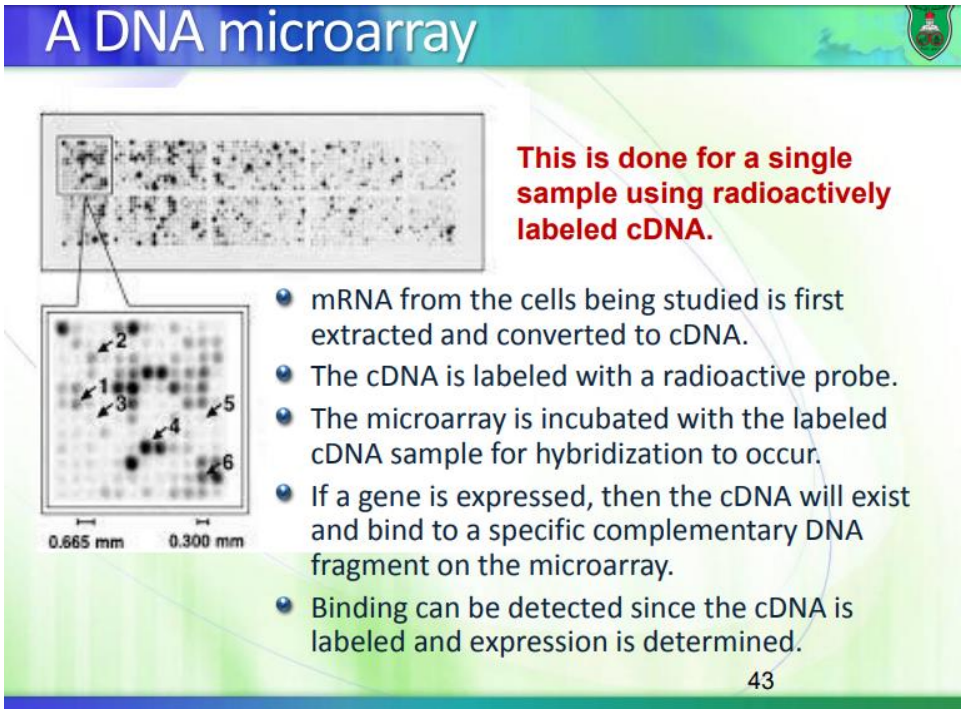
Array: Array: sequence of anything

Microarray = chip (contains thousand of "spots")



Spots contains probes of certain genes, and there are different probe molecules of one particular gene

eg. One spot contains cluster of probes molecules for x, another for y .



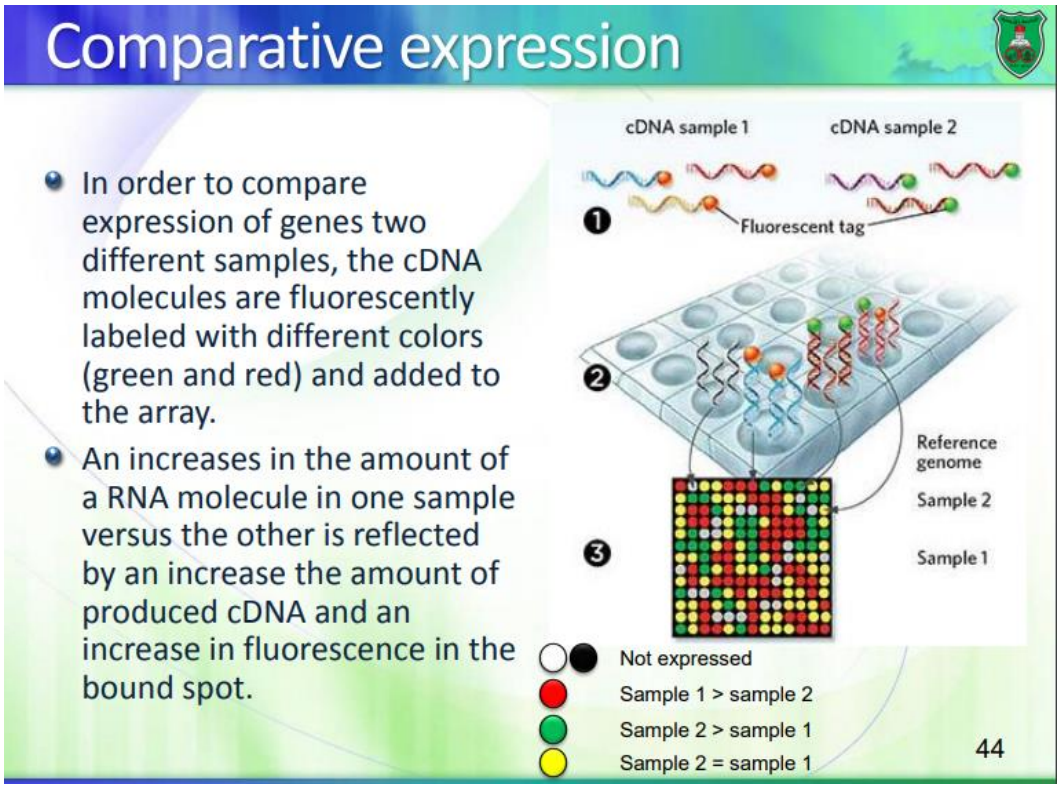
Notes:

*)Black dot: A spot that contain certain gene
*)Number of mRNA molecules = Number of cDNA
*)when the spot color is strong (more black) it gives more signal that means there are more cDNA so there is more messenger RNA, that means the gene is strongly expressed
*) We can tell the relative expression of genes at a certain time point by looking at the intensity of all signals

Steps:

1.extract mRNA and convert it to cDNA
2. label cDNA molecule (in this experiment we used Radioactive probe, but it isn't used now anymore)
3. Throw all the cDNA molecules on the microarray so each cdna will bind to its probe
4. If a gene is expressed, the cDNA will exist and bind to a specific complementary DNA fragment on the microarray
5.Binding can be detected since the cDNA is labeled and expression is determined

---

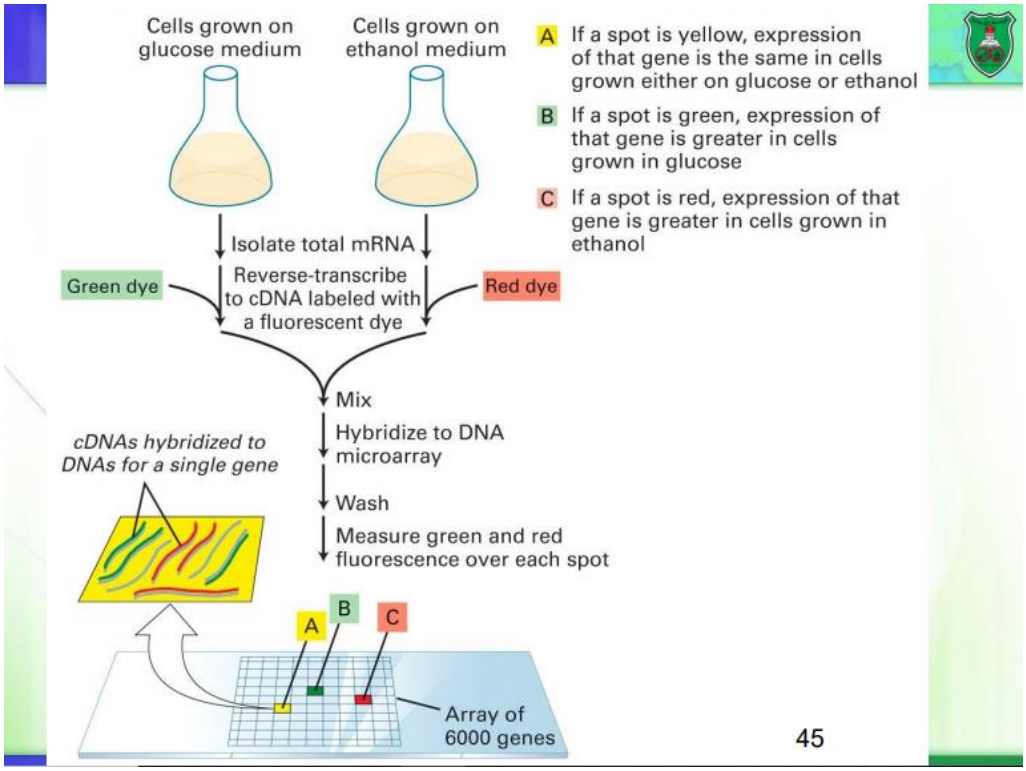


To be more accurate, we have to measure expression, not doing relatively.

*)comparative expression: comparing expression of all genes into samples at the same time

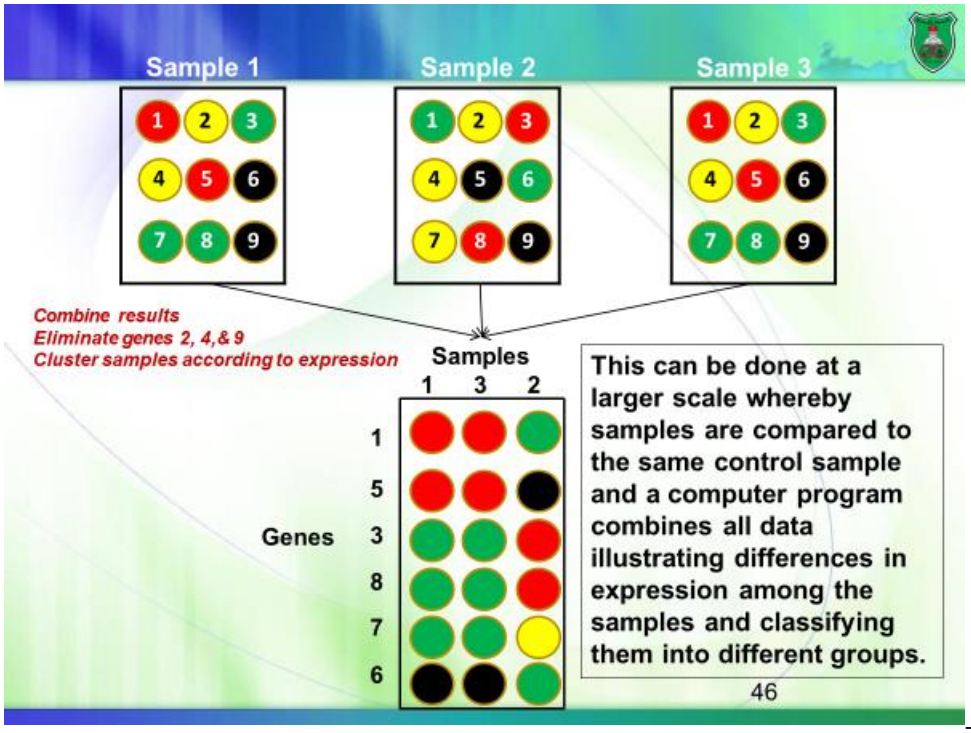
Eg. Sample 1 : controlled cells
Sample 2 : cells exposed to a certain drug, cancer cells(to decide which genes are important of transportations of control healthy cells to cancer cell), cells exposed to certain toxic (to see the changes in genes after exposing to this drug and that tells you how genes are important to eliminate this drug)
*)note: we used fluorescent here not radioactive elements
steps in comparative expression are almost the same with the previous steps, but here we used two different genes with two different fluorescent tag, each gene will bind in a different amount at each spot, that's why there is a variation in colours

---



*) Slide 45 is an example of comparative expression, Dr. Mamoun read it only

---



The trick here is not only doing the experiment itself, but the bioinformatics, that is the analysis of the results, no person would be able to look at all expressions of all genes and analyzing the spots, so we use computer.

*) the computer can do thousands of samples, so we have a lot of data points

*) The main advantage of using computer: to **CLUSTER** samples together, computer groups diseases that look similar histologically to different groups according to their molecular profile

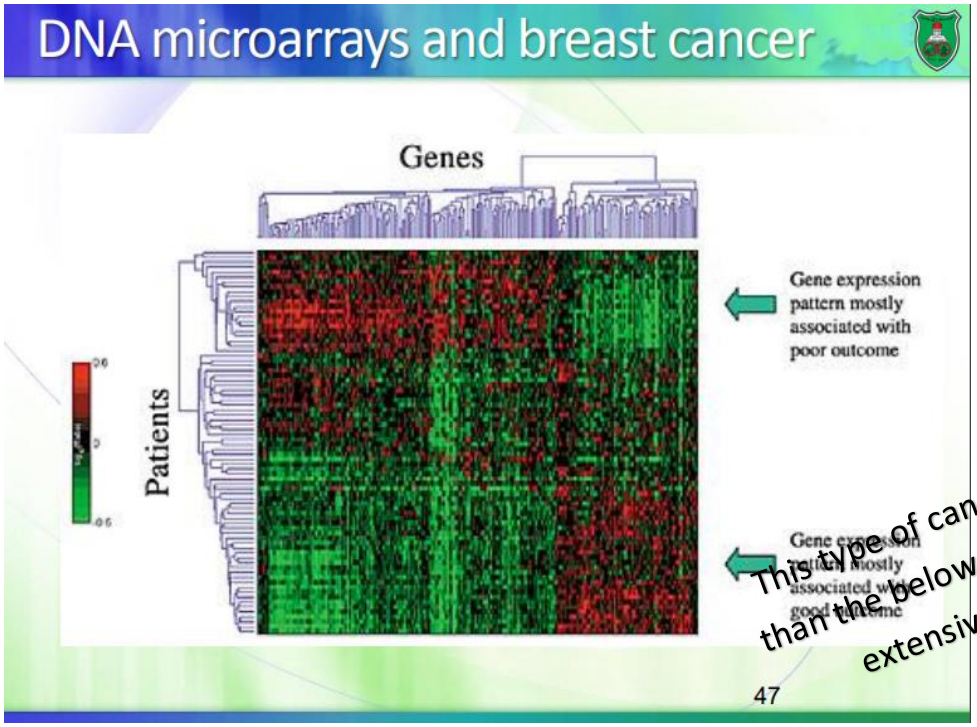
*)e.g lets say that the three samples are three patients of liver cancer, they contains different mutations that affect in different way, because the **molecular profile** for each one is different

lets say: green is for cancer, red is for normal, yellow when they are equal, and black when there is no expression

*) we didn't take spot number 9 for example, because in all samples it wasn't expressed, they are all the same, so it's not informative, also we didn't take gene number 4 and 2…….

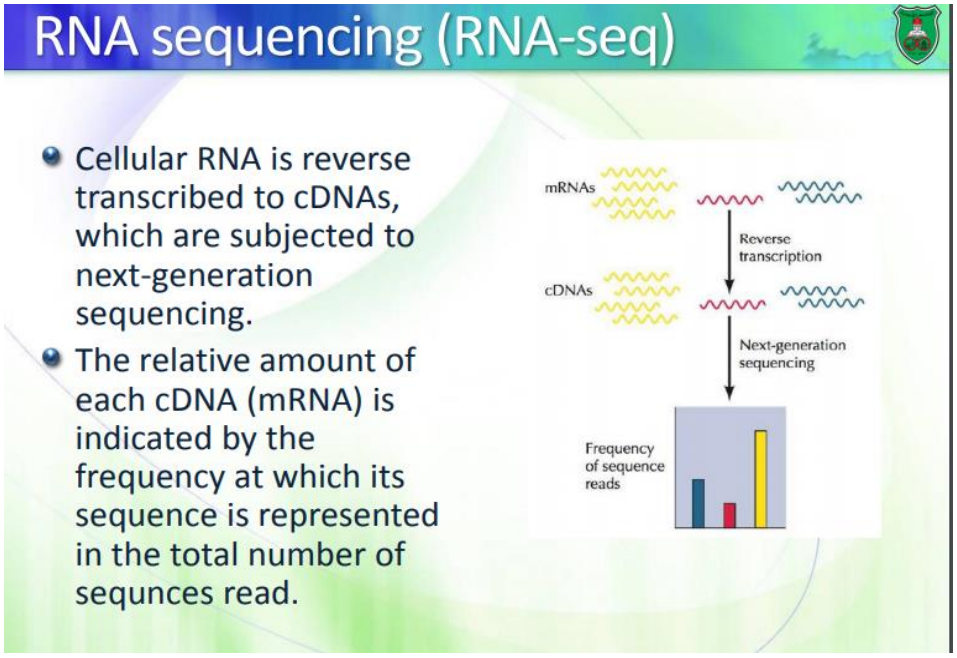
*)sample 1 and three are very similar, that's why the computer had put them beside each other

*) benefit of this process: to know how to treat with each sample



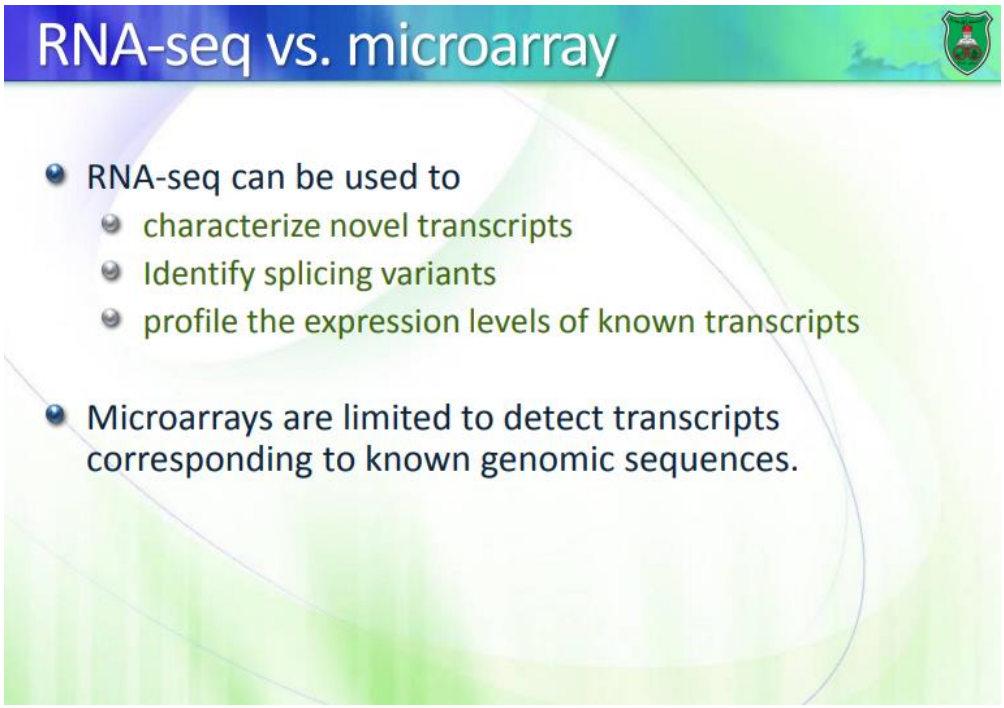
*)This is a real example of many samples of comparative expression, containing many of samples, it's used as a history to help detecting the type of cancer for a new patient to give him the suitable treatment

---



*) here we make sequencing of RNA, and use it for quantitative purposes, so we took all messenger RNA, and then convert them to cDNA, after that we sequence all of them, as a result; we can know how many time the cDNA are sequenced so we can indicate the quantity of the mRNA
*) yellow cDNA molecules in the previous example are more sequenced more than the blue one, which is sequenced more than the red one.

*) There is more yellow cDNA and mRNA than the blue, which are more than the red ones, so the gene of the yellow is more active than the blue, which is more active than the red.



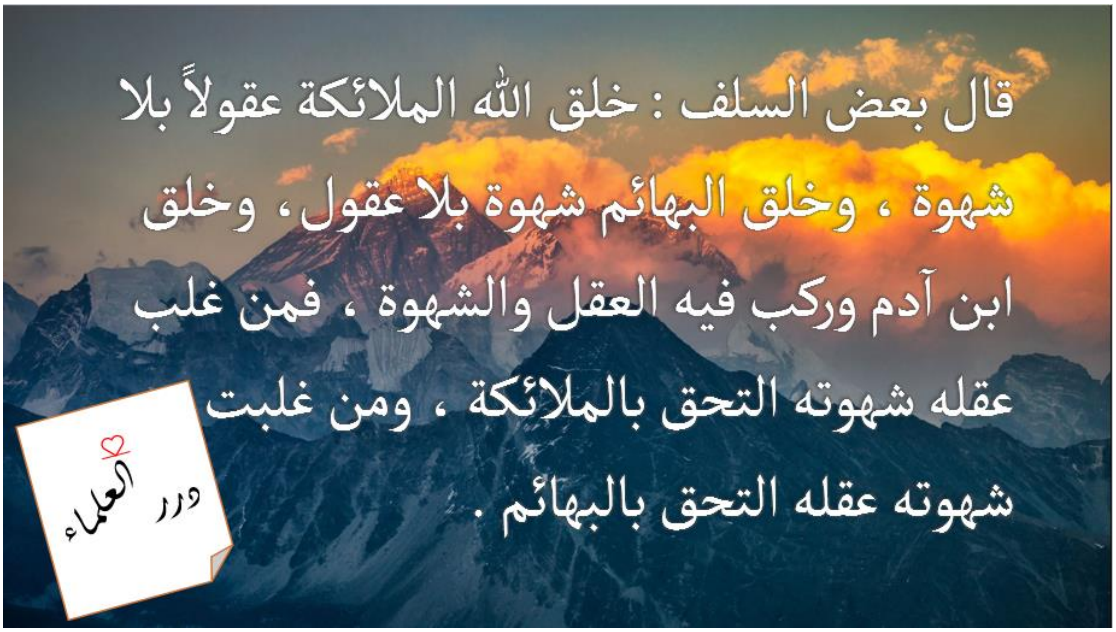
*) There are many advantages in RNA-sequencing in addition to know the quantities of cDNA and mRNA, we can know the sequences as well, and this is important because we can identify novel transcrips

*)remember with microarrays we know what the probes are that are placed in every single spots so we already have a knowledge of the sequence of the genes, but with RNA-seq, we sequence everything even if we know or not, and that allow me to identify novel genes

*) before 6 years scientist sequenced all RNA molecules, they compared it to the data base of the genome project at DNA sequence, they found that 70% of the human genome is transcribed

*) in fact, we can identify splicing variant of the same gene with microarray but hardly

---



SHORT QUIZ

TEST BANK

Q1) which of the following is found in cDNA?

A) The template for the mRNA that is examined

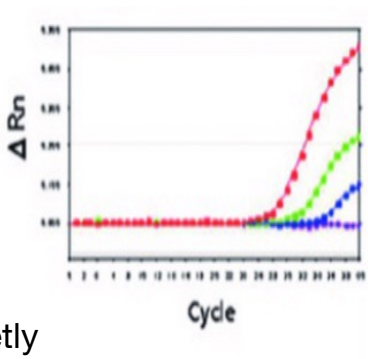
B) The template's complementary sequence

C) The mRNA that is examined

D) All the above

E) A + B

Q2) Based on this graph which represents RT-pCR , which of the following is false:



A) Housekeeping genes were definetly used in this experiment for validation

B) the blue labeled sample had the least amount of starting mRNA

C) The number of cycles for the red sample clearly outnumbers The other samples

D) premature detection was prevented by lack of machine Sensitivity

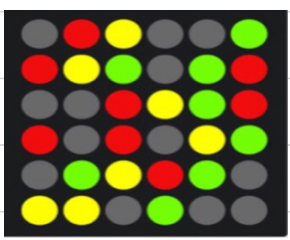
Q3) Which of the following is unique to RNA sequencing ?

A) It allows for the quantification of RNA expression.

B) It allows the study of novel trascripts in both human and nonhuman specimens .

C) It can identify disease by recognizing overly or underly Expressed genes .

Q4) based on the following microarray which was made with a healthy (green) and cancer (red) cells answer the following:



1) what do the yellow labeled probes represent ?

A) Genes that are non affected by cancer

B) Genes that are equally expressed between these cells

C) probes that were not linked with any genes

D) all the above

E) A + B

Q5) In a certain RT-PCR experiment that had three samples and was only sensitive enough to detect droplets after they hit the 9,000 mark which sample would be detected first and after how many cycles ?

(S.1=500 droplets ,S.2=250 droplets ,S.3=450 droplets)

A) s.2 after after 6 cycles .

B) s.1 after 5 cycles .

C) s.2 after 5 cycles.

D) s.1 and s.3 simultaneously after 5 cycles.

ANSWERS

Q1	Q2	Q3	Q4	Q5
E	C	B	E	D