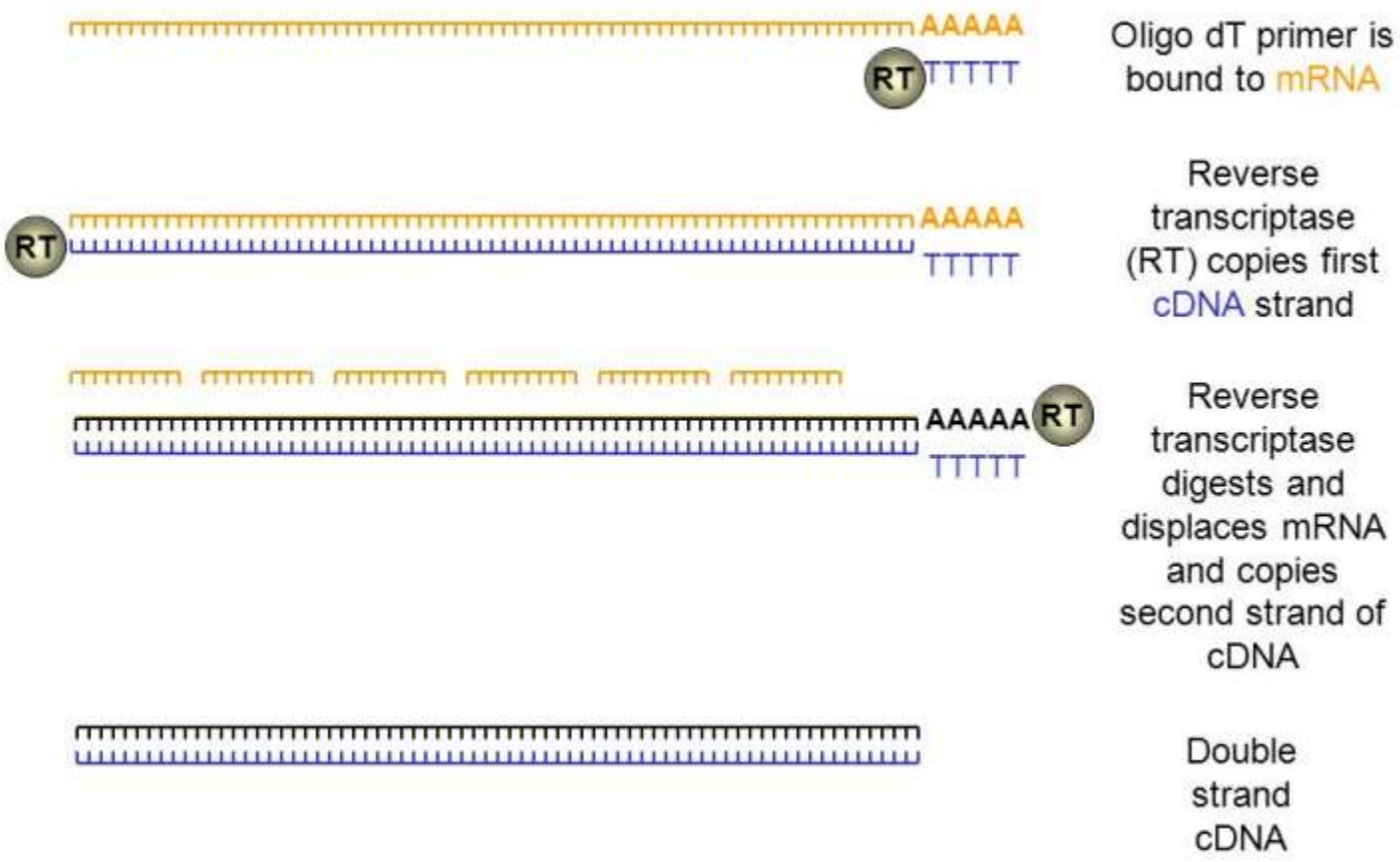
MED-HUB

BIOCHEMISTRY

Analysis of gene expression

Analysis of gene expression and RNA levels

- Techniques leaded to analysis of gene expression via measuring RNA levels:
  - **Basic methods** (northern blotting - in situ hybridization)
  - **Advanced methods** (real-time PCR - DNA microarray)
  - **Very advanced methods** (RNA-seq)
- To do real-time PCR for RNA, DNA microarray or RNA-seq, firstly, we should **make a complementary DNA (cDNA)**

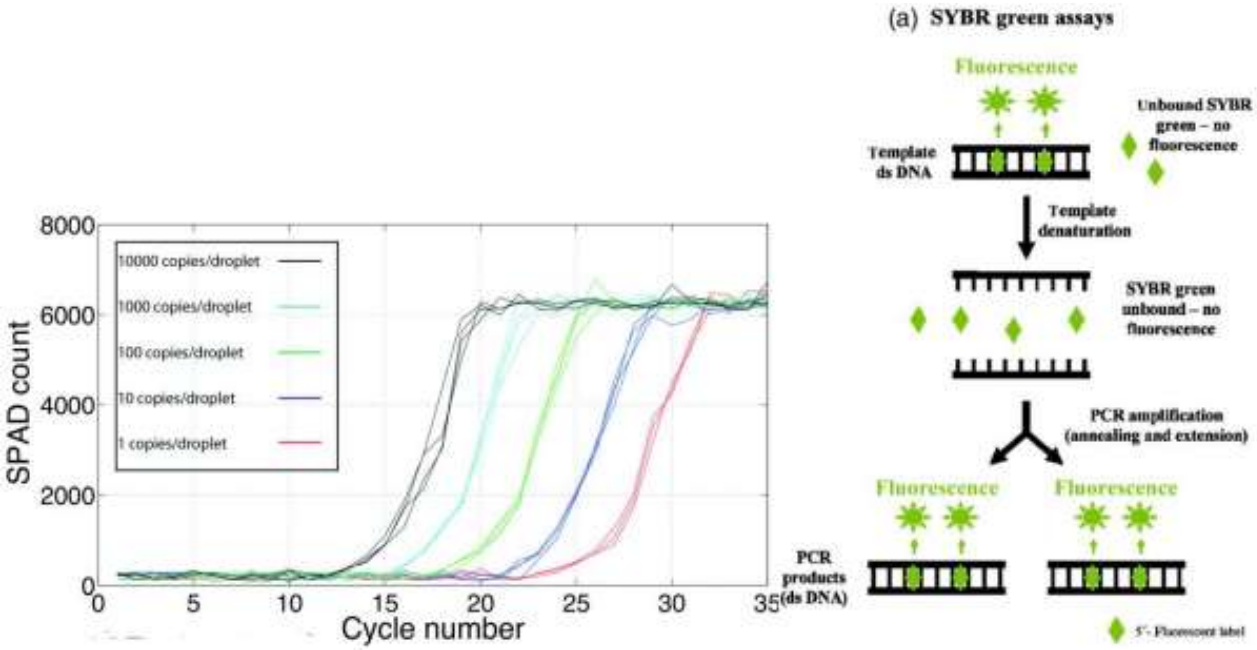


- A primer (oligo-deoxy T/ poly-T primer) complementary to poly-A tail is added to mRNA and then **reverse transcriptase (RT)** binds to it.

  NOTE: Each mature mRNA ends with poly-A tail.
- **Reverse transcriptase** makes a complementary strand to mRNA and copies first cDNA strand by using mRNA as a template.
- **Reverse transcriptase** degrades the mRNA molecule and synthesizes the second strand, now we have double-stranded DNA (cDNA).

Real-time qPCR of mRNA

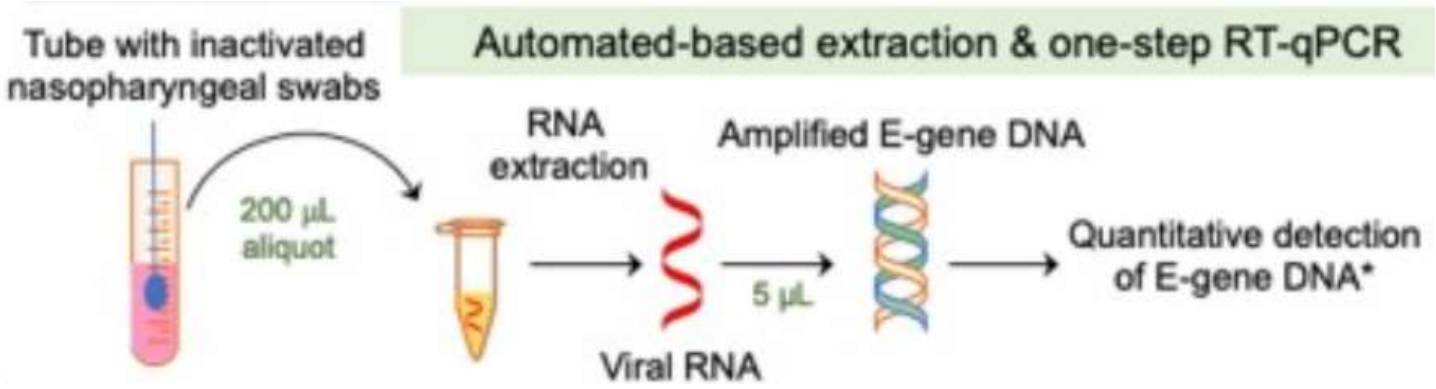
- This technique is done by **converting RNA into cDNA**, followed by **PCR** in the presence of SYBR green.
- The ratio between mRNA and cDNA is 1:1, if there are 10 copies of mRNA, you make 10 copies of cDNA.
- By qPCR, cDNA is amplified using primers and the amount of DNA can be quantified according to when the signal is detected.
- *The higher amount of RNA* >> *The higher amount of cDNA* >> *The earlier it is detected*.



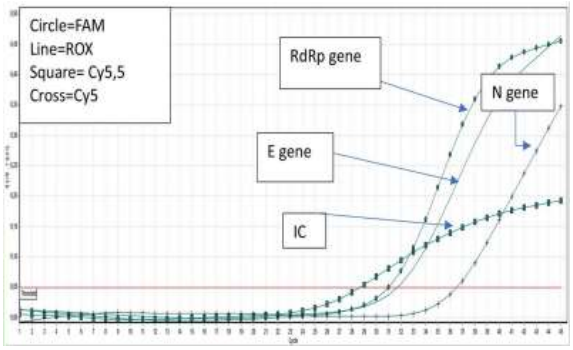
- **Housekeeping gene**: A gene that is expressed by cells all the time and its level is not different according to the environment that cells live in such as actin, tubulin which makes microtubules **(constant expression)**.
- If we want to do qPCR of RNA to 2 samples and compare the gene expression in them, we need **a control**, so 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.**

- We take the 2 RNA samples and convert RNA into cDNA, then we amplify the **housekeeping gene** (actin, tubulin,…) in both samples and **expression shouldn't change**, so we have an **overlap between the two plots**. *If the housekeeping gene expression is different between samples, this means that the amount of total RNA in samples isn't the same.*
- **NOTE**: IL1-beta is a cytokine produced by inflammatory immune cells.
- You can see above that **the detection of cDNA in one sample is earlier than the other**, so **the expression in this sample is higher than the other**, and **we are sure of our results because the expression of the housekeeping gene is the same**.

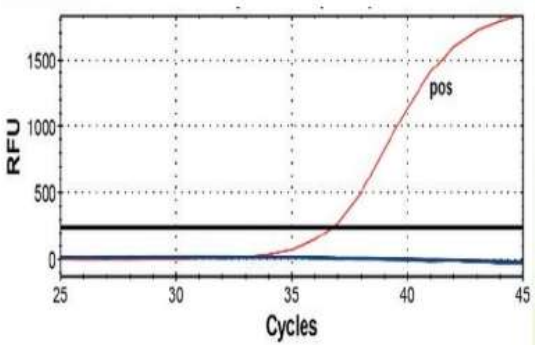
Detection of SARS-Co-2



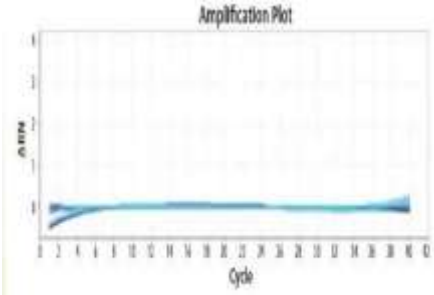
- We use this technique to know if someone is infected with corona virus or not, by converting the viral RNA into cDNA using **reverse transcriptase** and then we do **real-time PCR** for it.
- We amplify corona virus genes, like **E-gene**, which is a specific gene for SARS-CoV-2; actually we amplify 3 genes to confirm that it's SARS-CoV-2.
- We also use **an internal control**, the IC, which is an amplification of human RNA, to make sure that the sample is collected properly.



- **In the first picture**, the signal is detected for IC which means that the sample is collected properly, and the 3 genes of SARS-CoV-2 are detected, so the patient result is **positive**.



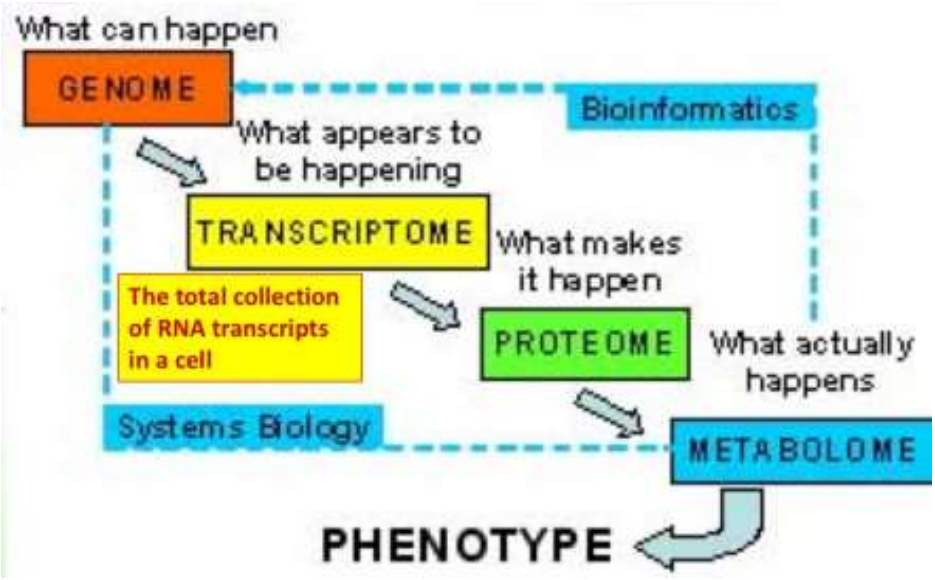
- **In the second picture**, we can see the control and the viral genes aren't detected, so the result is **negative**.



- **In the third picture**, the control is not detected, meaning that there is a technical issue, so we can't conclude anything**, the sample is invalid**.

The science of –omics

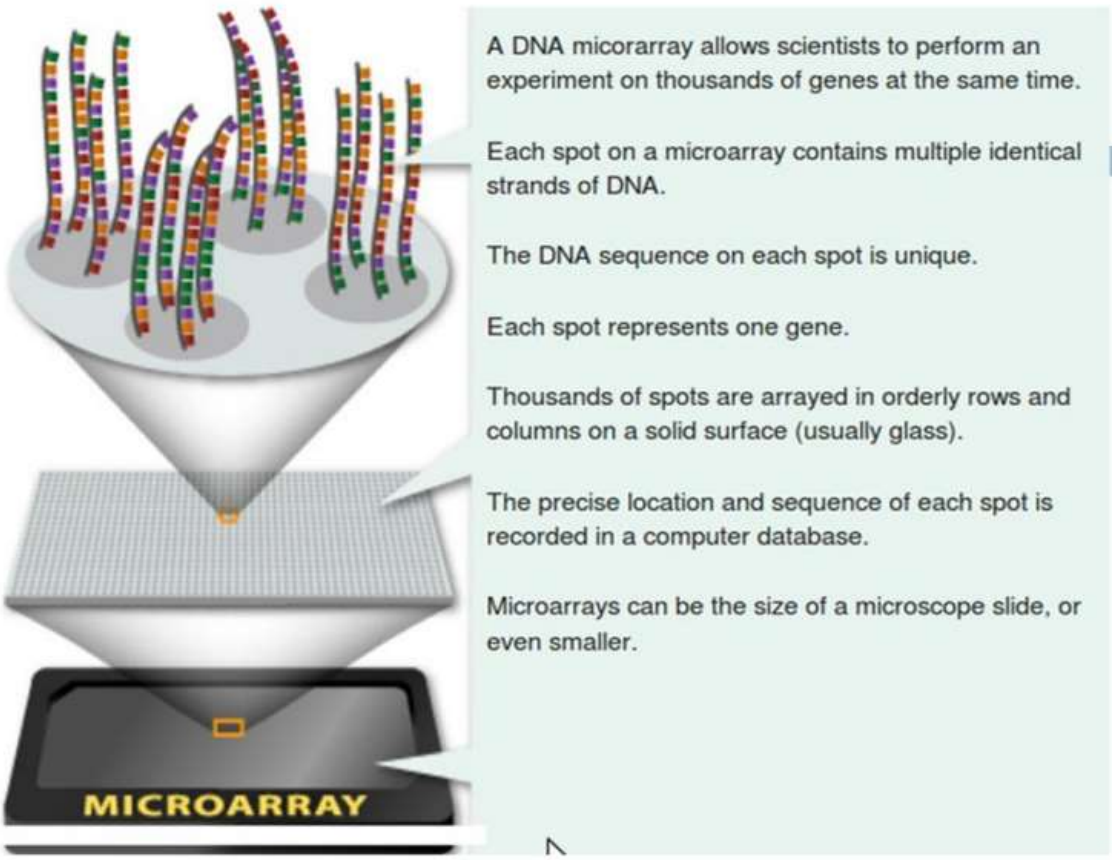
- **Genome** is the whole DNA in a cell, and the science that studies it is called **genomics.**
- **Transcriptome** is the total collection of RNA transcripts in a cell, and the science that studies it is called **transcriptomics**.
- **Proteome** is the total collection of proteins in a cell, and the science that studies it is called **proteomics**.
- **Metabolome** is the collection of metabolites –biochemical reactions' products- in a cell, and the science that studies it is called **metabolomics**.



- We are interested in transcriptomics, which uses **DNA microarray**.

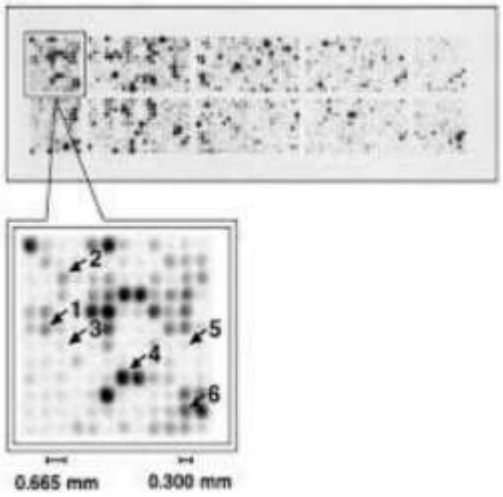
DNA microarray

- **DNA microarrays**: Solid surfaces (glass microscope slides or chips) spotted with up to tens of thousands of DNA fragments (**probes**) in an area the size of a fingernail.
- **The exact sequence and position of every DNA fragment on the array is known**.



- This technique allows 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.

- This picture represents a glass slide, and we zoom in to see the spots; **each spot contains identical probes for a specific gene**.



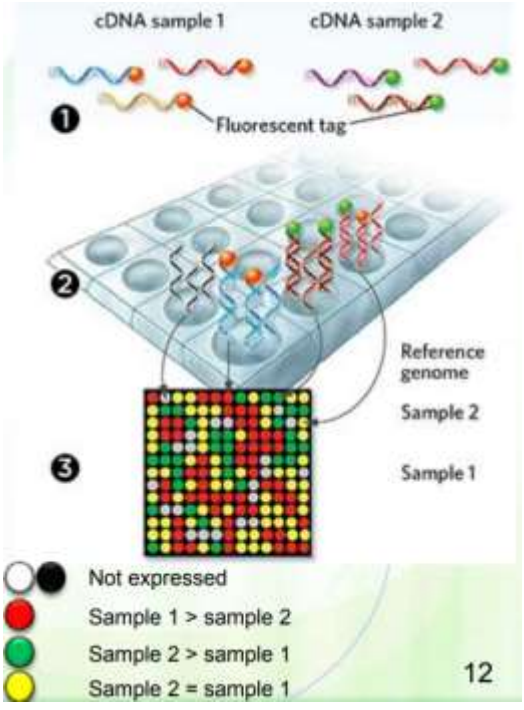
- **The process**:
- mRNA from the cells being studied is first extracted and **converted to cDNA** using **RT**.
- **The cDNA is labeled with a radioactive probe**.
- The microarray is incubated with the labeled cDNA sample for hybridization to occur.
- If a gene is expressed, then the cDNA will exist and bind to a specific complementary DNA fragment on the microarray.
- Binding can be detected since the cDNA is labeled and expression is determined.

  - *The more cDNA molecules bound to the probes in a certain spot, we will get stronger signal*, **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**.
  - By comparing **the intensity of the signal** between the spots you can tell which gene is expressed more than the other genes.

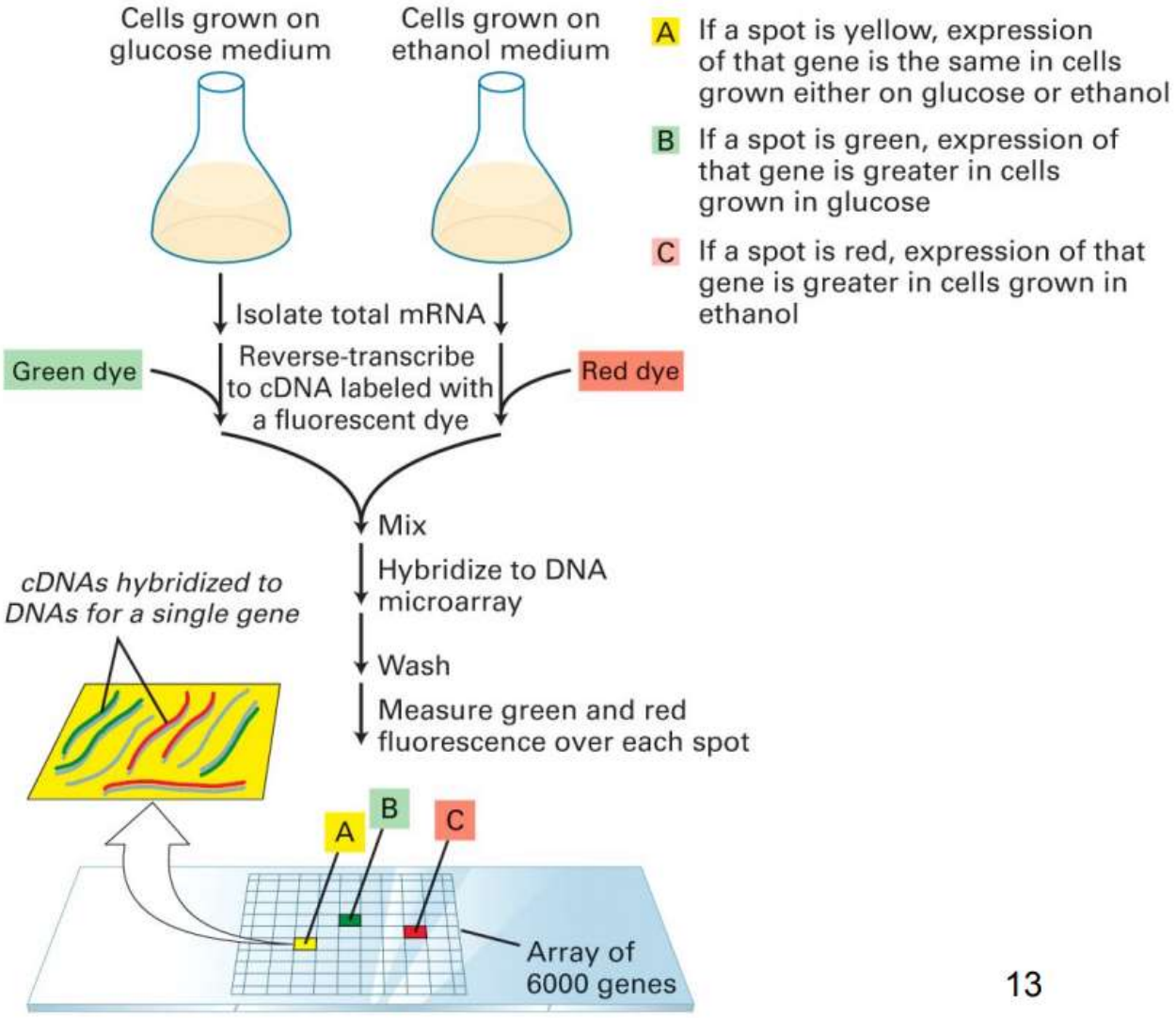
  ➢ DNA microarray tells us about **gene activity of all genes in a cell** at a certain time.

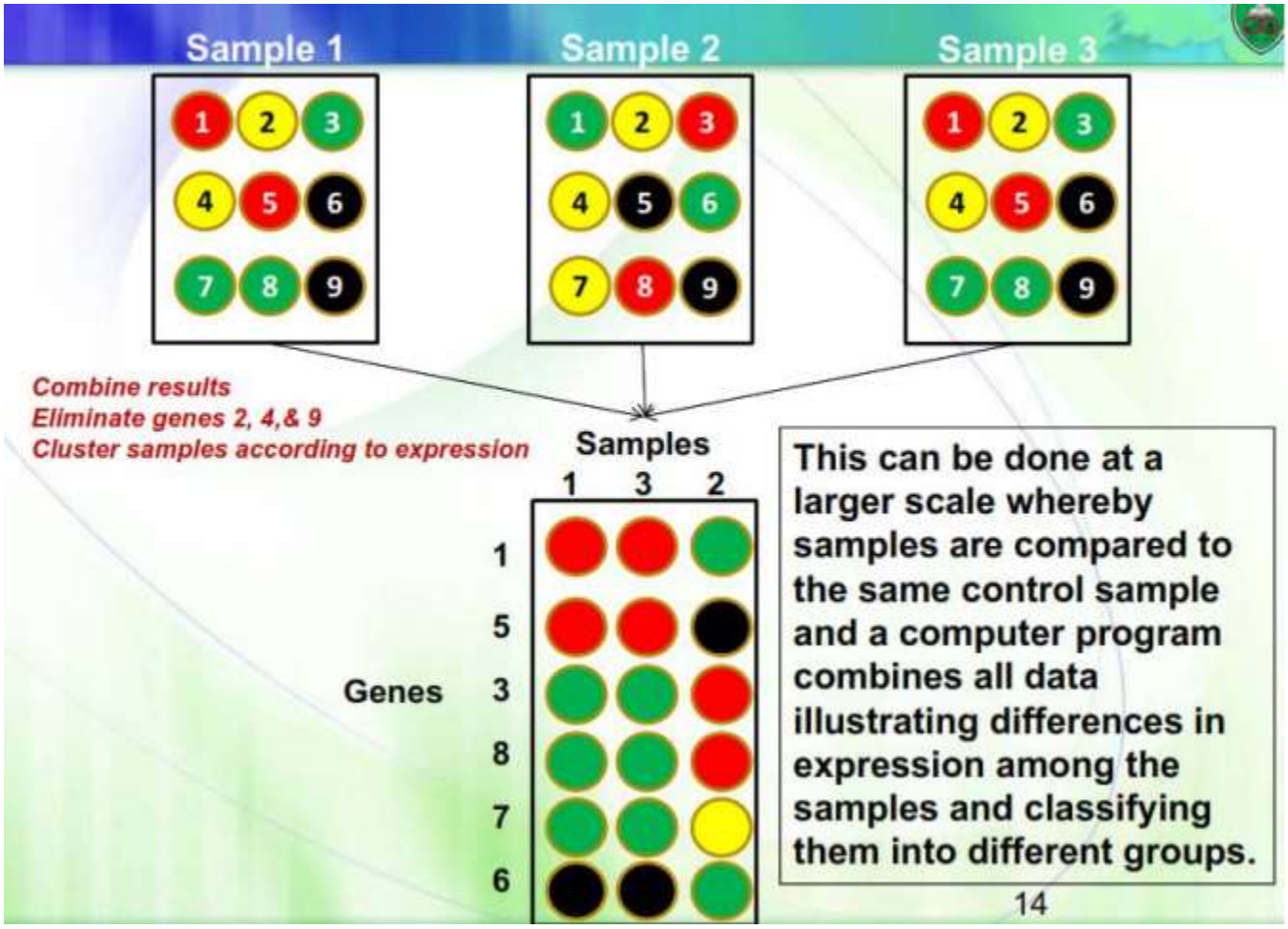
Comparative expression

- We want to **compare gene expression between 2 samples** (normal cells-cancer cells), to see what genes are differentially expressed between different samples and these genes could tell us something about the cause of cancer, for example the gene can be an oncogene (a gene that can cause cancer) or a tumor suppressor gene (a gene that prevents tumor formation).
- With radioactivity, we can't compare gene expression in 2 different samples, as it has no distinct color.
- We take RNA molecules from 2 different samples and, using **RT**, 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 >> at the spot, red cDNA will bind with the probes more than the green cDNA >> there will be **red fluorescence**. (and vice versa)
  - **If there are equal amounts** of red and cDNA >> at the spot, it will generate equal amounts of red and green fluorescence and it will give us a **yellow color**, and it means that 2 samples have the same level of gene expression.
  - If the gene is not expressed in both samples (no mRNA or cDNA), there will be no signal.

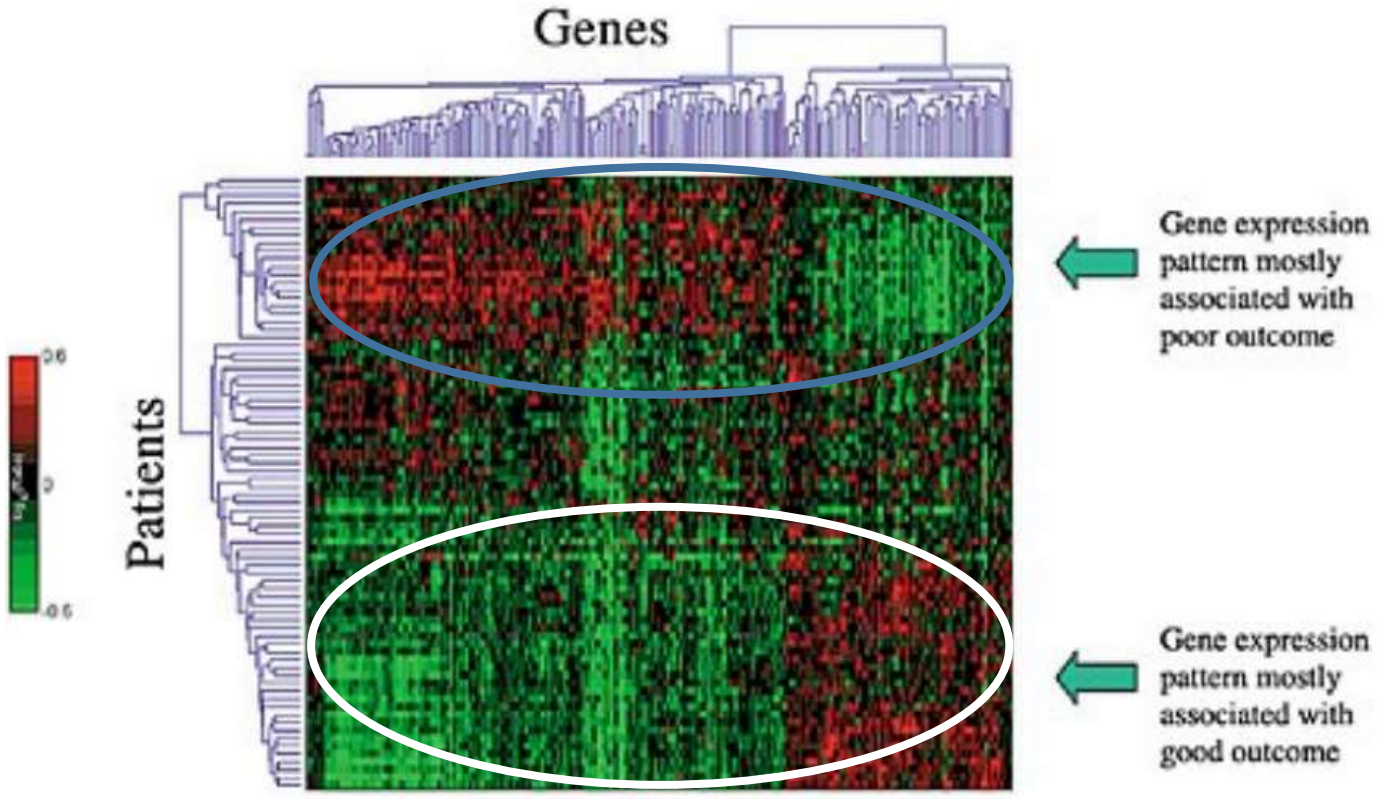


- Here is an example:





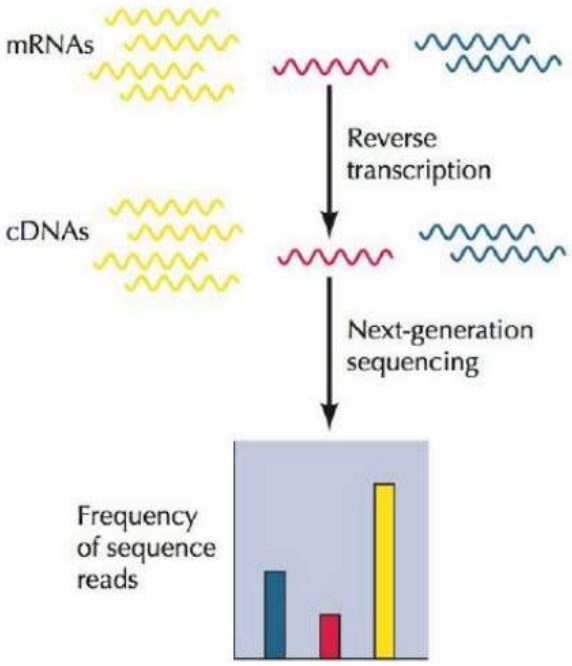
- We use informatics, and computer makes a comparison between the 3 samples and it **eliminates similar genes** (2-4-9), because they aren't informative, and it puts similar samples together to ease the comparison.
- **For gene 1**, there is an **overexpression** in samples 1 and 3 vs a common control, and it is **under-expressed** in sample 2, or we can say it's **down regulated** in sample 2.
- **For gene 5**, the gene is **not expressed** in sample 2, and so on.



- This picture represents a real example of results.
- We have samples taken from different patients with **breast cancer**, and we analyze the **gene expression patterns** for a lot of genes.
- **Patients in the blue circle** have **over expression** of the genes at the left and **under-expression** of genes at the right, and these patients have in general **poor outcome**.
- **Patients in the white circle** have **under-expression** of the genes at the left and **overexpression** of genes at the right, and these patients have in general **good outcome**.
  - **Prognosis (outcome) can be linked with gene expression!**
- Next time, we take a sample from a patient and we analyze the gene expression pattern and put it with the first or second group, then **predict patient's prognosis** and it can also **affect treatment plan**.

RNA sequencing (RNA-seq)

- **RNA molecules** are isolated from a cell and **reverse transcribed to cDNAs**, which are subjected to **next-generation sequencing**.
- **Remember**: the ratio between mRNA and cDNA is 1:1, if there are 10 copies of mRNA, you make 10 copies of cDNA.
- 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.



- **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):**

It can be used to:

- **characterize novel transcripts (discover new genes)**; microarrays are limited to detect transcripts corresponding to known genomic sequences because we design probes that are specific to cDNA.
- **Identify splicing variants** of the same gene; while in microarray we can have 2 variants bound to the same cluster, but here we sequence everything.
- **profile the expression levels of all transcripts**, while microarray is limited to known genes.

- Using RNA-seq, we knew that 75% of the human genome is transcribed.

Past paper

- Using a fluorescent-based DNA microarray, a computer-generated yellow color means
    a. Expression is higher in one sample versus the other
    b. There is expression but no cDNA in the sample
    c. There is an equal expression in both samples
    d. There is failed binding of cDNA to the attached probes
    e. There is no expression in either sample

- The best technique to discover the expression of a novel gene is
    a. Protein tagging
    b. DNA sequencing
    c. DNA microarray
    d. PCR
    e. RNA sequencing

- Using radioactive-based DNA microarray, comparative expression cannot be done on the same slide (the solid platform) because
    a. Radioactivity has a low level of detection
    b. Radioactivity has no distinct color
    c. There is a lower hybridization capability of glass slides
    d. The amount of probes on the slide is very little to handle two samples
    e. Using two labeled samples means high radioactivity and this is unhealthy

Answers: # DONE BY MED-HUB

c – e – b