University of Jordan
Faculty of Medicine

GENETICS & Molecular Biology

Number: 14

Done By: إيمان أحمد العموش

Corrected By: Mohammad Qussay Al-Sabbagh

Subject: DNA Library ,RT-PCR & microarrays

Doctor:

Price:

Date:

BE A WARRIOR, NOT A WORRIER.

In the last lecture we talked about how **we can study gene expression**. We talked about **Northern**, we talked about **In situ hybridization** and we'll continue that but with **some techniques used in the laboratory**.

Let's start ! (Slide 330)

The question is what is a **library** ?

A library is basically a place you go to. It has collection of books, you search for a certain book, you go to the right floor, right floor, then you take the book that you want and you go home with it. And you do whatever you wanted to.

We can apply the same concept to **DNA libraries**. Basically they are clones of bacterial cells that have plasmid inside. And these bacterial cells containing the same plasmid. These clones are saved/placed in a freezer, so they are frozen in order to study them. And the freezer is full of bacterial clones. Each one of them contains a certain plasmid and each plasmid contains a certain DNA fragment. So every time you want to study a certain gene or a certain DNA fragment all you have to do is searching in the data base.

For example: I searched in the data base for a specific DNA fragment. Oh I want this bacterial fragment that is present in freezer number 5, shelf number 3, row number whatever, box number whatever, location number whatever. I take the frozen vial, I take the bacterial cells, I grow them. And now I have the fragment that I want to study from. That's a DNA library, a smart concept.

There are basically two different libraries that can be created :

1)Genomic DNA library.

2)cDNA library.

Genomic DNA library (Slide 331)

The Genomic DNA library. It is basically that I take the genome of a cell (the whole content of chromosomes), then:

1) I cut it into different smaller pieces.

2) I insert these pieces into a plasmid.

3) Put a plasmid inside the bacteria.

Now each bacterial cell has a certain segment of a genomic DNA. If I take these bacterial cells and I look at the DNA, the DNA can be part of a gene. It can be an exon, an intron or a promoter. It can contain coding or non coding genes. Whatever, its genomic DNA. And what's important is that I can create the same genomic DNA library from any cell in the body ! , Because they all contain the same genome. The chromosomes in my skin cells are exactly the same as in my muscle cells, as in nerve cells and so on.

so that's the first type of a DNA library.

cDNA library (Slide 332-333)

The cDNA library. **c** stands for complementary. cDNA represents the mature mRNA. So basically it only contains the exons (no entrons), and it can contain the 5' untranslated region and the 3' untranslated region, but it contains the mature RNA molecule.

The way it is made is the following:

We extract all RNA molecules form the cell. Containing mRNA, rRNA, tRNA. What is the thing that exists in mRNA that doesn't exist in rRNA or tRNA ?

There is something that is added to all mRNA molecules and it doesn't exist in rRNA or tRNA, it is the poly A tail is only added to mRNA molecules and not added to other RNA molecules.

So I can simply isolate the mRNA molecules by taking the molecules that have the poly A tail and then I convert the mRNA to DNA, double stranded DNA, using an enzyme known as reverse transcriptase. And this enzyme is present in retro viruses which are viruses whose genome is RNA not DNA, because what they do is they infect our cells with their RNA molecules and then within it they have the reverse transcriptase that converts the RNA into DNA and DNA integrate into our DNA and that's it. That how it should work.

So that's the way by which we can create a cDNA library by taking advantage of the poly A tail of the mRNA molecules. Good ? :P good !

Remember that the cDNA is a representative of the mature mRNA that contains the exons as well as the 5' untranslated region and the 3' untranslated region

RT-PCR of mRNA (334)

It is a technique that is used widely in the laboratories. known as real time PCR. It takes advantage of this technique that we talked about before, the SYBER green.

This dye (SYBER green) binds to double stranded DNA only, so the more double stranded DNA the more fluorescence SYBER green will give us. So we can quantify how much RNA molecules we have in a cell, simply by taking a mRNA, we convert it to cDNA, and we amplify the cDNA.

Now for every molecule of RNA we have one molecule of cDNA. If you have 1000 molecules of a certain mRNA well have 1000 molecules of that cDNA, so its 1:1 ratio.

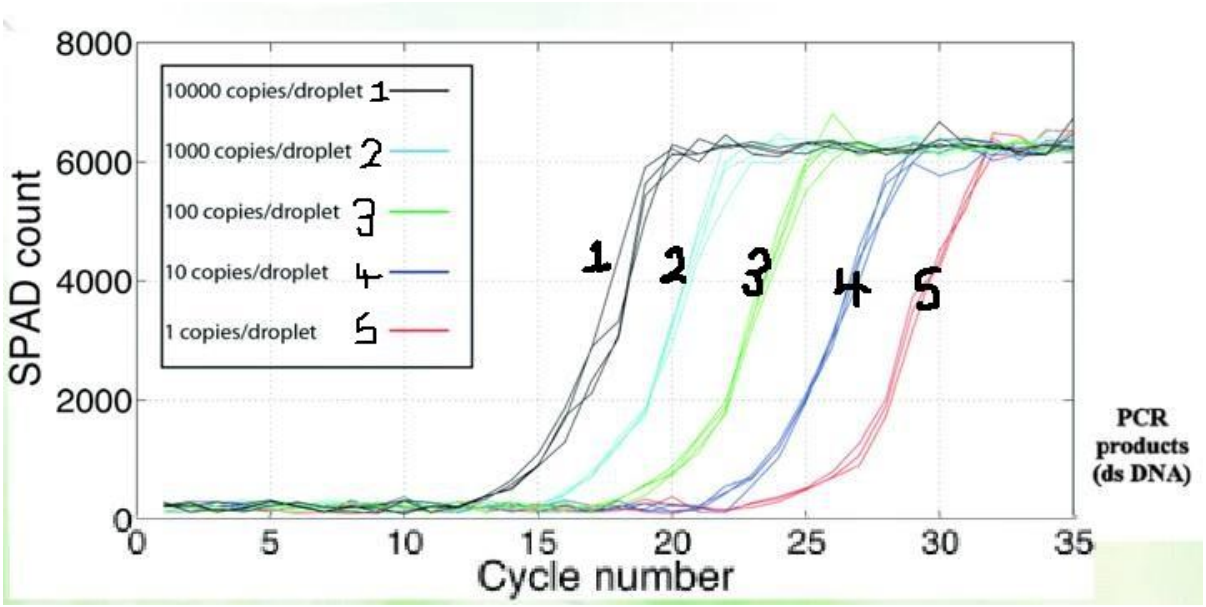
So when u amplify the DNA or cDNA the more cDNA u have the larger the signal, which means that the correlation is the higher the expression the more the mRNA you are having in the cell, the more cDNA you can synthesize, and when you do the PCR, the higher the signal, so that is the sequence of events, so u can quantify how much u have of a certain RNA molecule and that gives you a reflection of the gene expression, the activity of the gene, using these RT- PCR as well as SYBER green.

Now the thing is that as lower as the cycle the cDNA is detected in, it means the higher the amount of the RNA molecules. The lower the signal the higher the RNA molecules, that is the higher the cDNA molecules that is the more the gene expressed.

A student asks the doctor to re-explain the main idea and he explain it here:

Just like doing quantitative PCR using RT-PCR, same techniques and same idea. The more the gene is active, the more mRNA it makes, the more cDNA you synthesize because its 1:1 relation, the higher the signal

or in other words the lower the cycle that this cDNA is detected in. So the more the DNA, of course you are making more and more of it early on, so it is detected early on.



For the following you need to take a look at the figure above.

Sample number 1 contains more cDNA, initially more mRNA than sample number 4, and sample number 2 contains more cDNA than sample number 5, because it detects the signal at an earlier cycle. The signal starts to appear at cycle 16 for sample number 2, but the signal is detected at cycle 29 for sample number 5, which means there is more cDNA in sample number 2.

So this is RT-PCR, it's another way of quantifying how much message you have, how much gene is expressed. In other words, with northern blotting we look at the intensity of the signal but it's really not that accurate. RT-PCR is much more accurate.

Look at the box inside the figure, it tells you that sample 1 has 10000 copies that are detected early on when they are amplified. Sample 5 has 1 copy that you can see it but you see it later on in cycle 29 or 30 whatever.

The science of-omics (Slide 335)

Scientists came up with a science and they called it genomics, which means the study of the whole genome at the same time. It's the age of technology, it's the age of speed.

In the old days they used to study one enzyme, one gene, one aspect of a gene and that's it ! Someone who would study a promoter, another person would study exon 5 and so on. But these days you can do a lot more, so they came up with a science and they called it genomics. They said we can now study RNA molecules, we can study gene expression and they called this science transcriptomics by studying the transcriptome, that's by studying mRNA molecules, the level the amount of mRNA.

Scientists loved that word and they said: well, why won't we study all the proteins of the cell ? and they called it proteomics which is the study of all proteins in the cell collectively. They are studying protein – protein interaction, protein expression as a whole not only one single protein.

Studying the transcriptome/DNA microarrays (Slide 336/Slide 337/Slide 338)

How can we study the transcriptome as a whole ?

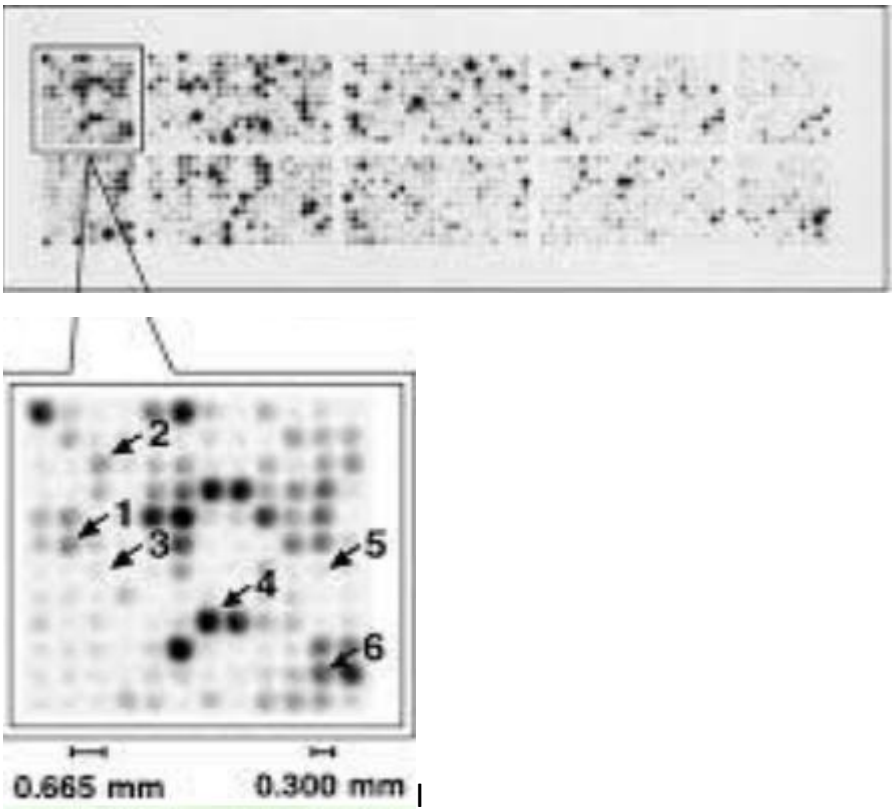
Doing northern blotting, we can analyse expression of 5 genes, but its time consuming, what about RT-PCR ?

We can do it, its not bad, its multiplexing meaning that you can study expression of several genomes at the same time, because in this you have one sample divided into 5 tubes, you add primers here you add primers here, for specific gene, so you can do multiple analysis at the same time, but again limited number of genes that you can study with RT-PCR, In Situ Hybridization and northern blotting.

So they said "no no no, I want to study the expression of all genes at the same time" . This is the science of transcriptomics and its really really exciting. So they came up with a technology and they called it DNA microarray.

Array is مصفوفة, it basically several genes lining up. Micro because its small. DNA because each one would contain a certain DNA for certain gene. It looks like a chip, it's the size of part of a finger! Its really small like 1cm^2. But it contains probes for thousands of genes so you can analyse thousands of genes at the same time in one single shot. Amazing technology.

A DNA microarray (Slide 339)



If you amplify this, you can see that its really an array, meaning that it has wells and spots, each spot contains a probe for a single gene, so you know from the array for example that position number 1 is the probe for gene x, position 2 is the probe for gene y, position 3 is the probe for gene z. So you can have the 22000 probes for 22000 genes all lining out right here, and only you have to add a cDNA , and that's how It works !

So there are several methods, one of them is the following:

1)You take all the RNA molecules

2)You separate only mRNA molecules using the <u>poly A tail</u>

3)You convert the mRNA into cDNA ,So the amount of cDNA represents the amount of mRNA

4)You take the cDNA and you label it with a certain colour and you add the chip to the array.

Then each cDNA will bring to a certain position depending on the probe , if it hybridizes with the probe or not. So if there is a signal it means that there is a cDNA that binds to that probe. If there is no signal, it means there is no cDNA, which means that the gene is not expressed which means the following: **No gene activity→ No transcription → no mRNA → no cDNA → no signal cause there is nothing to bind in that position.**

So if you look at the position 4 for example you can conclude several things. at this position there is a signal. It means that this is a probe binds to the cDNA so there is expression. In portion 5 there is no signal, for that particular gene there is no expression in that cell.

Compare position 4 to position 1, you find that position 4 is more intense why ? because there is more cDNA which means these is more mRNA which means there is more gene expression. So expression of this gene is higher than the expression of that gene.

So from microarrays you can :

1)Determine if a gene is on or off.

2)Determine the extent by which the gene is expressed. (is it high expressed or not expressed or low expressed).

So that's the micro array ☺

Comparative expression (Slide 340/Slide 341)

We can do a lot more with microarrays , we can do something that is really beautiful.

We can do comparative analysis. It means we can compare expression of RNA of genes from two different cells together on the same array. And we do this by taking the mRNA form the cell and taking the mRNA from another cell. Let's say that we have normal cell and we have diseased

cell. Let's say that we have control cell and we have a cell treated with a certain drug, and we want to know what genes this drug affects.

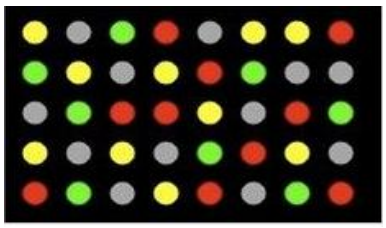
We take the mRNA from both samples, we convert it to cDNA, we label the cDNA with green colour in a sample and with red colour in the other sample. So let's say you have one gene that's equally expressed, 1000 RNA molecule here and 1000 RNA molecule there, 1000 cDNA, 1000 cDNA, 1000 would be labelled with green and 1000 would labelled with red.

When you put them on the array , the green signal would be equal to the red signal , that tells you that the expression is equal.

Let's say that a particular gene is not expressed in both cells, no mRNA in both, no cDNA in both , you add nothing to the array and a particular spot there is not binding because there is no cDNA, there is no signal so you know that there is no expression from this cell or that cell.

Let's say that you have a gene that is highly expressed when you add the drug. let's say normally you have 10 RNA molecules, 10 cDNA molecules, and after treatment you have 10000 RNA molecules , 10000 cDNA molecules, you add them both, one of them will give us higher signal, you will have more red than green, so you know that relatively this gene is expressed after the addition of the drug.

Similarly, let's say that expression of a gene is suppressed after the addition of a drug, it means that the other signal would be higher than the other one. So we can do comparative gene expression analysis using DNA microarrays.



You would get something like this:
meaning if you got a black or grey spot it means there is no signal, no expression from both cells.

If you have yellow it means that expression is equal in both cells.

If you have red it means that you have induction and if you have green it means that there is suppression on gene expression. That's what it tells

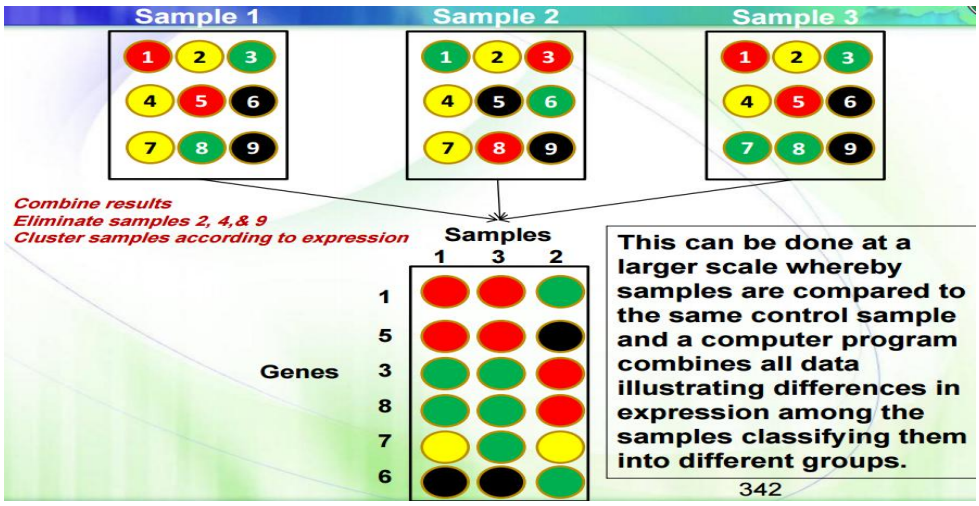
us, so what does that mean, it means that if I want to study the effect of this drug on these cells, i just need to focus on the genes that are with changed expressions.

Note : Every spot represents a gene, I am looking at changes in gene expression, so the gene is normal , there nothing with it, I am looking at its expression not at its mutation. We are comparing gene expression in the same cell after treatment of a certain drug. So every spot represents a gene so I know that expression of this gene is not altered, this gene is not expressed in any of the cells before or after treatment, this gene is down regulated, this gene is up regulated and so on. So that's the information that I get, changes on the gene expression.

In the previous slide (DNA microarray) we see black colour ! if I got black signal from a cell and a black signal from another I cannot compare expression between two cells at the same time, I just have to add cDNA from one cell to the array in this case, I take another array and I add the cDNA from the another sample. But here I am doing the same array I am adding cDNA from this cell and from that cell, so am doing the comparison on one array so they give me different colours.

Question : can't we compare according to the density ? yes of course, we can . but this is more beautiful Since it's colourful. In the DNA microarray we use radioactivity which is dangerous, and here we are using fluorescence which is safe, in DNA microarray we cannot quantify how much signal there is, its just low intensity or high intensity , but here we can measure in real numbers the amount of the signal.

Slide 342



Let's say you want to do comparison between different cancer cells. So you take mRNA from normal cells, and you take mRNA from cancer cells of patient one, then from patient two and from patient three.

Then what you do is that you take each mRNA and you convert it to cDNA and you combine the normal cDNA with the cDNA of patient one, and you added it to a microarray, you do the same with patient two, you take the cDNA from patient two and you add it to the control to the normal cell and you add it to its microarray. And so on.

So here we have the same reference; the same healthy gene expression, and you are comparing patient one with the healthy, patient two with the healthy and patient three with the healthy and so on, and you get different signals, each one of each sample will give you certain signals.

Now on the array we have thousands of probes, thousands of spots, its microns so it's really hard for me to as individual to look at each position and compare it, rather it's the computer that does the homework for me, so the computer :

1)Combines all of this data together.

2)It compares the expression from the different samples.

3)It groups the samples together according to expression.

So it says for example: look here, we have three samples, I combine the data and here we have 9 different genes.

The computer tells me that sample one and sample three are similar to each other, why ? because you look at the expression of gene 1; its high, but in sample 2 its low, and gene five is high in these samples (1 and 3), but it's not expressed in sample two.

Look at genes 3 and 4, they are down regulated in the samples (1 and 3), and they are up regulated in sample two. Look at gene 7, its different here so let's puts it on the side. Gene 6 is not expressed in samples one and three, but it's down regulated in sample two.

So, basically the computer tells me that samples one and three are similar to each other, sample two is different.

So the question is how I can differentiate between samples one and three from sample two ? I look at the expression of gene 1 that tells me

something, or I look at the expression of genes 3 and 8 because they're different. So it gives me idea, it tells me that genes 1, 3 and 8 are important for these samples, they do something, they give me a hint about these genes.

So what we have learned is that :

1)We can classify or cluster samples according to gene expression.

2)These microarray tells me which genes are important for samples one and three.

So you look at samples one and three, you go back to the log book and you look at the identification of these samples. Wow, samples one and three are aggressive cancers, advanced tumour cells, but sample two represents a benign tumour.

So that tells me something, it could be gene 1 or gene 5 that have driven these two cells to become really aggressive and probably suppression of expression of genes 3 and 8 also make the cells more aggressive, so maybe these are tumour suppressor genes because they are down regulated and these are oncogenes because they are up regulated, so you can tell a lot of things from this array.

Slide 343

This is a microarray done on tens of cancer samples for different patients, and these are hundreds of genes. The computer tells me that these genes cluster these patients together, these patients are similar to each other according to the genes which are up regulated and the genes that are down regulated in these patients.

So, I can classify the tumour according to molecular expression rather than histology, and what they do these days is that they look at these genes, it's the opposite, they are up regulated in these patients and down regulated in these patients and that tells me something as well, that also these genes are important so we use them to classify the tumours according to gene expression, and this can tell me about the importance and the significance of these genes in the progression of the

disease, in the diagnosis of the disease, and in the terminate prognosis that is how sever the disease is.

And that's what they do in cancer nowadays, in cancer biology they stop, they haven't really stop, they still look at the tissue under the microscope and say this is basal carcinoma, this is whatever, but by doing molecular expression they can more accurately determine and tell me something about the diagnosis of the cancer, the type of the cancer and the prognosis.

Look at these patients for example, they all have cancer but these are the patients that have died early on, and those are the patients that have survived for a long time. So what are the factors that have determined that these patients would die and these patients would survive ? it's the expression of certain genes.

So if a patient comes to the clinic, I tell him let me look at the gene expression, oh its good prognosis you have a good chance to survive a longer time, or you know what, let me give you aggressive chemotherapy because your cancer doesn't look good. That's help us determine something about cancer, molecular classification of diseases.

Sometimes you can be surprised by the way because some of these genes can be oncogenes and tumour suppressor genes, and sometimes it can be metabolic genes that are up regulated or down regulated. Its not necessary that this gene is a tumour suppressor gene or an oncogene, but it tells us something about this gene that can be important for cancer to survive, like a metabolic enzyme for example.

Translation

Now , we are going to talk about the **last** topic in molecular biology !

Translation :is the process which cellular ribosomes produce proteins from mRNAs .

We have 3 main types of RNAs involved in translation process , which are : rRNA , mRNA and tRNA

tRNA

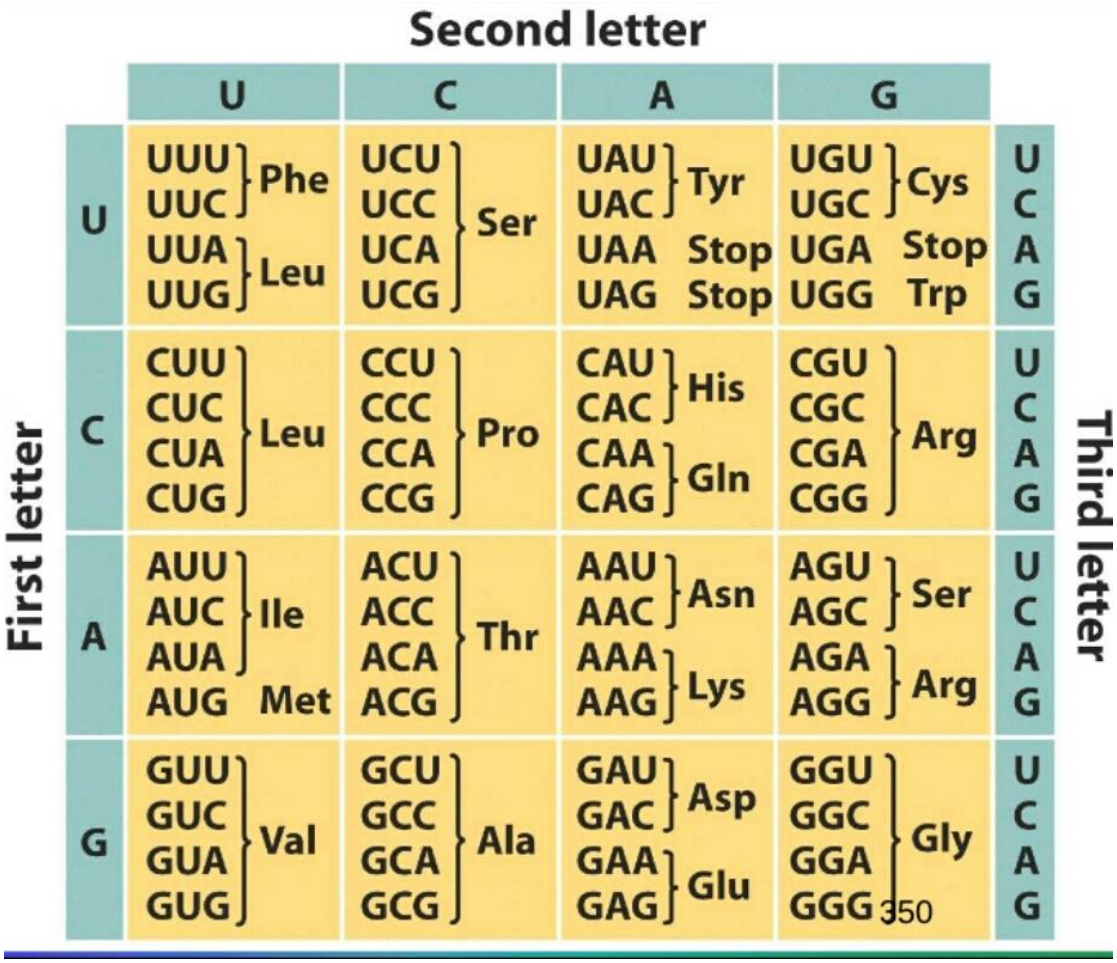
tRNA : is a very special short RNA molecules (80 bases long), that carry one amino acid on it's 3' terminus , all tRNA have CCA at the end , and the amino acid is linked there to adenosine .

Charged tRNA : is a tRNA molecule carries amino acid , this is done by an enzyme called Aminoacyl-tRNA synthetases .

anti-codon : part of tRNA that interacts with mRNA during protein synthesis .

note : The anti-codon is <u>antiparallel</u> to the Condon , that means if I have a 5'-AUU-3' codon , it's anti-codon will be 3'-UAA-5' , or more accurate <u>5'-AAU-3'</u>

The genetic code



You are not supposed to memorize this table , it will be written in the exam .

But basically it is the arrangement of the nucleotides on the mRNA .. for example UCG codes for serine , if you look for AAA → lys

So if the doctor asked you ,about the anticodon of UGC for example , it will be GCA (from 5' to 3')

What is interesting about genetic codon that we have three end codons , which are : UAA , UAG and UGA .. which are responsible for termination of translation . Notice that these codons are not universal , since that some bacteria have different terminating codons .

degenrate codon : degenrate means (منحط) , degenrate codon is the codon that could be changed , but still give you the same amino acid , but this Wobble base pairing is only true when the change occur at the 3rd base on the mRNA of the codon .what happens is that tRNA binds to the first and second bases specifically , but when it comes to the 3rd one it is not that specific ..

We take an advantage from this , if we got a mutation on the 3rd position , it wont change that much on the protein .

For example , AUU , AUC and AUA code for the same amino acid , and even if we got a mutation that changed the codon to AUG , we still in the safe side , still have similar amino acid with minimal change on the protein .

Moreover , even if the 1st or 2nd bases are changed , it will give me similar amino acid , so we still protected ..

The end

في رحلتك الطّويلة هذه، ستُدرك يوماً أنّك في عزّ تبرعُمك كنت مخطئًا في ترتيب أبجديّة عُمرك؛ كان عليكَ أنْ تكون اللام والعين والنون، حتى الياءُ فيك، كان عليك أن تكونها أنت بقوّة وأنانية ! فأنت الحرف الأول والأخير.. وما سواك كلمات، أنتَ حرفها والقلم فيها فؤادك .إيمان أحمد العموش