Genetics & molecular biology

◉ Sheet ○ Slide

Number:	12
Done by:	Feras Suleiman
Corrected by:	Bayan Abusheikha
Doctor:	Mamoun Ahram

How do transcription factors regulate gene expression?

In bacteria: 1-Lac operon, we said that the RNA polymerase attaches to the promoter region and if the Lac repressor was present, then the RNA polymerase doesn't move.But if it wasn't there,then it opens up the promoter and will have the open promoter complex and transcription will start.

2-CAP, if its bound upstream of the RNA polymerase binding site, it interacts with the RNA polymerase.It stabilizes the interaction between the RNA polymerase and the DNA, making transcription efficient

---

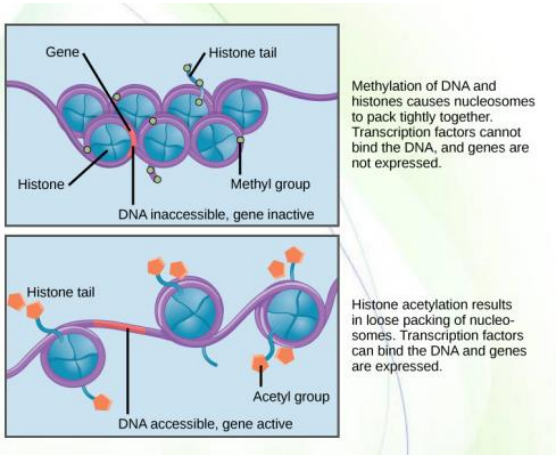
In eukaryotic cells: (*things are more complex*), we have chromatin (*combination between DNA* and histones). If we look at two people's gene expression rates, and one if them had quicker gene transcription.It's suggested that the person with the slower gene transcription has a SNP in the promoter, a SNP in the polyadenylation signal, or he has a mutation and the RNA is not stable. Thenif the DNA sequence, RNA sequence and the protein sequence are checked,then we'd find out there is no mutation whatsoever.

Then how was there a change in the gene expression?

- It is due to what's called e**pigenetic control** or **epigenetic regulation,** basically it is a higher level of gene expression (as epi means above or in addition to). Scientists have just recently started knowing the real meaning behind epigenetics. So by epigenetics, DNA can be changed in a way that gene expression can be regulated. This can happen by :

- Chromatin packaging
- Chemical modification of histones
- Chemical modification of DNA

Modulation of chromosomal structure:



In karyotyping (genetics),you will notice that on chromosomes there are dark and light bands. These bands indicate that we have *euchromatin* and *heterochromatin*. *Heterochromatin* is when the DNA is packaged tightly around histones. And the light bands indicate *euchromatin*. Active genes will be in the euchromatin region, because the DNA is exposed so DNA sequences are clear to the transcription factors where they can bind.

While in the heterochromatin the DNA sequence is hidden. So by epigenetics we can regulate gene expression by hiding the DNA or exposing it.

- How is DNA packed?

As the DNA polymerase reads the DNA, the chromatin remodeling factors remove the histones allowing the DNA polymerase to synthesize. In addition to removing histones, chromatin remodeling factors can reposition histones,(*they can bind to the DNA and push the histones aside*) making the sequence visible for transcription factors or attracting all of the other molecules we discussed before like transcription factor 2D.

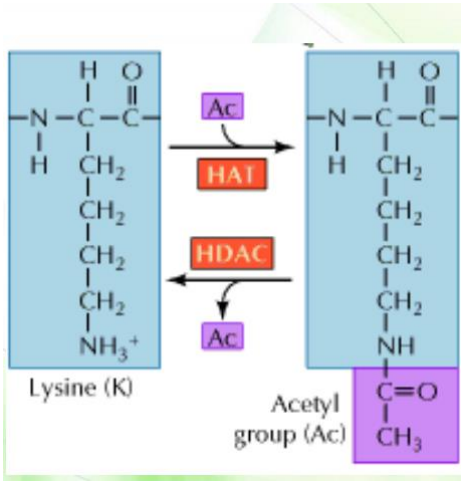
Note: We said before that transcription factor 2D (TF2D) is first to come bringing along RNA polymerase and others. How does it know where it should bind to the DNA?

1) As long as the DNA sequence is obvious (shown), it will bind
2) Transcription factor 2D is not only attracted by the DNA sequence but the DNA structure as well. The structure has a **bent** that attracts the transcription factor 2D.

This is part of what the activators and repressors do. They bind to the DNA and they remodel the DNA. Repressors remodel the DNA to be tightly bound to the histones, so they remodel it to heterochromatin (deactivating gene expression) and activators do the opposite.

We have different types of histone modification that takes place, some of which are *acetylation, methylation, phosphorylation, or sumoylation* (adding small protein called ubiquitin to histones). The non-coding RNAs to DNAs can recruit proteins that can modify histones.

- Acetylation:

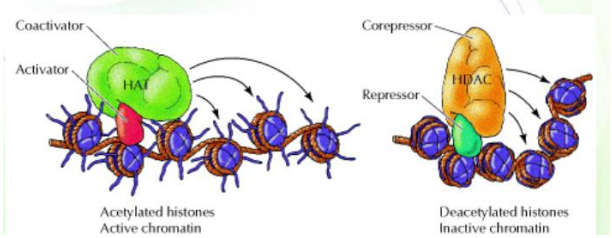


Histones are highly positively charged proteins because they contain amino acids (arginine and lysine) we are concerned with the lysine residue for now.

When the acetyl groups attaches to the lysine, the positive charge is **masked** and this results in weaker interaction between histone and DNA.

We have histones acetyl transferases and what they do is they add acetyl groups to the histones, hiding the positive charge and weakening the interactions between the histone and the DNA (loosening up the DNA which allows transcription factors to bind to the DNA).

On the other hand, we have histones deacetylases, they remove the acetyl groups, so they inactivate gene expression by packaging the DNA tightly creating heterochromatin. If we look at banding patterns of chromosome in DNA in different tissues, the banding patterns will be different because the chromosome needs to be expressed in one place more than the other(*a heterochromatin in one tissue might be a euchromatin in another tissue.*)

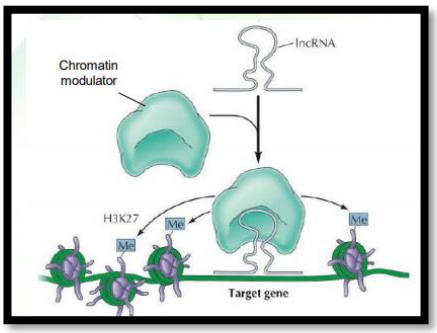


When TFIID binds to the DNA it has an *acetyle transferase* activity, it loosens up the interaction between histones and DNA, and as a result the sequence becomes clear to other transcription factors including

the RNA polymerase. And the repressors have deacetylases activity that makes the DNA tightly packed to histones.

For methylation and phosphorylation, it is a bit different. It depends on the histones (3,4, 2a and 2b) that gets modified, the lysine and the gene itself.(*But we don't have to know the details*).

- Non-coding RNA molecules



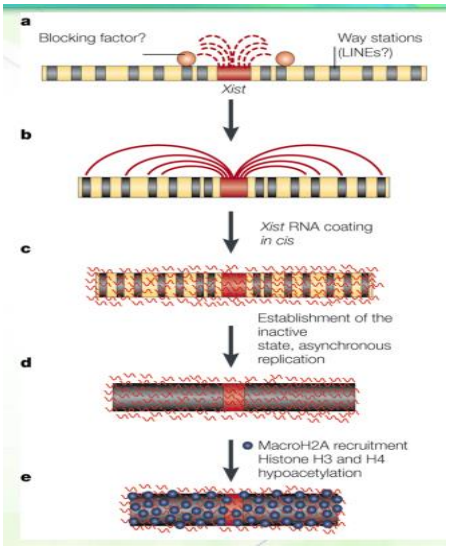
Scientists were happy they know the 21k genes, but now they found other 25k genes that code for non-coding RNA molecules, (*they don't become proteins but they are expressed*). There are so many types of these non-coding RNA molecules and are still being studied. Simply as the mRNA is being transcribed, the non-coding RNA molecules can bind to the mRNA and they can block synthesis, destabilize (maybe the stabilize mRNA we don't know), they stop gene expression and so on. So they have any of the three inhibitory effects. These non-coding RNA molecules can bind to the DNA itself producing a complex at certain sequences, inactivating gene expression. The idea is that they can bind to the DNA and attract the chromatin remodeling factors, histones chemical modifiers (acetylases and deacetylases) and this way they regulate gene expression.

An example of a non-coding RNA molecule and how can it regulate gene expression:

Females have two X chromosomes while males have one. During development one of the X chromosomes in females get inactivated. If we look in female organs for this chromosome we will find in some organs one X chromosome is activated while in others the other X chromosome is activated. This is called **mosaicism** and we will study it later on.

This inactivation of the X chromosome happens when non-coding molecules induce methylation of the chromosome, so the chromosome shrinks and becomes darkly stained (the barr body).

When we look at how it gets inactivated, there is a gene located on the X chromosome known as Xist. This gene gets expressed, producing non-coding RNA molecules, as it is being transcribed it binds to the whole chromosome, coating it, inducing proteins to **methylate** the chromosome.



However, even after the shrinkage there are still 70 genes that are active in both chromosomes, we know the function of some of them only but they need to be activated.

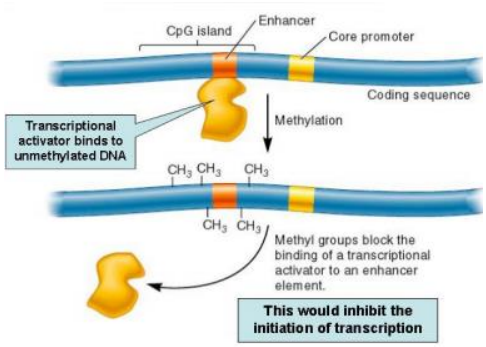
We have X linked diseases (X linked dominant and recessive),but being affected with the disease depends on the gene that gets mutated, because the gene might cause a dominantly inherited disease but females can escape this disease if the mutated gene is in the chromosome that gets inactivated (will not be expressed but the good one will).

At the same time the female can be effected by a recessively inherited disease if the ‘’good’’ gene was inactivated and the bad gene is expressed. So it depends on the gene, which chromosome gets activated or inactivated, and the number of cells.(it is all random.)

- DNA methylation

Cytosine can be methylated,(*we have methyl cytosine*). And if cytosine gets deaminated it becomes uracil, whereas when methyl cytosine get deaminated it becomes thymine.

The reason cytosine gets methylated in the first place is due to genes that have promoter regions that has what is known as CPG islands, this means a lot of Cs and Gs. If this CPG region gets methylated, we say that this gene is hypermethylated and the gene is inactivated.

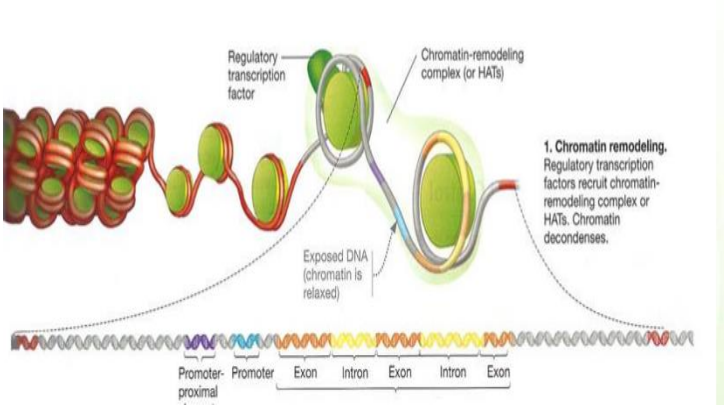


If the CpG islands are unmethylated (hypomethylated) the gene is active (transcriptions factors are bound and can activate gene expression)

So simply by not changing the DNA sequence but changing the methylation pattern in the promoter region you can simply change gene expression.

This methylation pattern is very important because it can effect another genetic phenomenon called genetic imprinting. Genetic imprinting means either the paternal gene or the maternal gene for certain genes can be activated. The methylation pattern can be inherited. In certain genes (not for X chromosome specifically but any chromosome) we all have to express the paternal gene and certain genes where we only express the maternal gene. If the expressed gene was the other gene we would have a disease. (*Will be discussed more with Dr. Bilal as it will be related to certain diseases*).

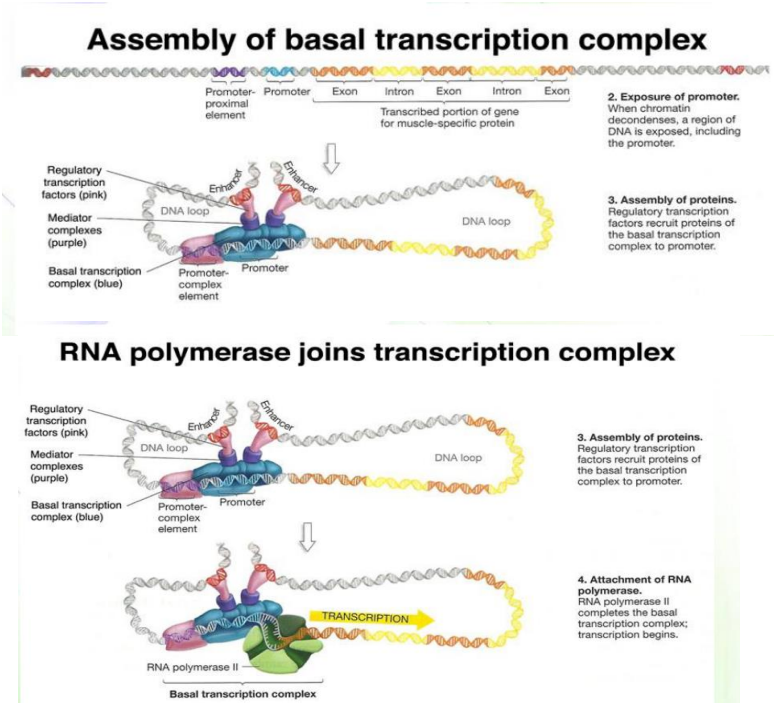
The chromatin remodeling factors expose the promoter region, attracting other proteins to interact with the DNA, one of these proteins is TFIID, modifying the histones.



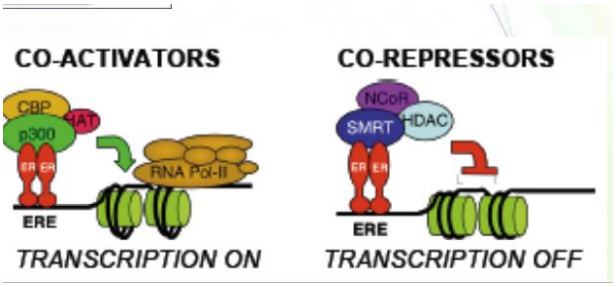
Then all of the other proteins are attracted to the promoter region and to the enhancer and interaction takes place between all these proteins then transcription starts.

- What are enhancers?

Certain DNA regulatory sequences that exist in DNA , they can be located far away from where the gene is (a thousand kbp distance), so the promoter and the promoter proximate element must be close to the gene but the enhancer can be far away because of the DNA feature (looping). DNA can loop around in different orientations stimulating interactions between all of the different proteins. If we change the place of the enhancer region it will still be function (if within the range). Even if we rotated it (what is infront becomes backwards) it will still be functional because the loop can turn as well. Even if we placed it in the gene in the intron region, it can still be functional.



As we said before about steroids nuclear receptors, that they are found in the cytoplasm, the hormone enters (estrogen, progesterone, etc..) by diffusing into the plasma membrane and they bind to a receptor, the receptors dimerizes,and the enter the nucleus. And they bind to the response element that is responsible for that receptor. And then the activators come (histone acetyltransferases) or the repressors (histones deacetylases) and they start to modify the promoter region, inducing attraction of other proteins to the promoter region activating or deactivating gene expression depending on the type of modification that takes place on the promoter region.



- Note :we have 70 genes on the X chromosome in females that are active on both copies of the chromosome, they of course know what the 70 genes are, but they don't know what is their function exactly.
- Note 2: Twins have the same genes, same DNA sequence, but their gene expression can be different. How? Be living a good healthy lifestyle, eating healthy, working out, etc. This is true to an extent, so don't blame your parent for your failure.

SLIDE: MOLECULAR BIOLOGY 8

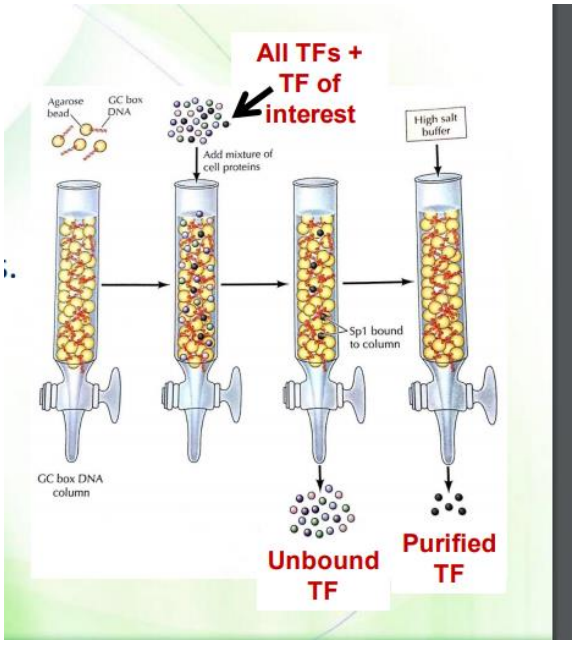
So in this lecture we are going to talk about a couple of techniques that are related to gene expression,we want to study the effect of that gene on the cell to know the function of the gene, the gene expression pattern (all of our cells contain the 21k genes we have) but is gene expressionis not identical. We have insulin in the pancreatic cells and we have neuro proteins in nerve cells. This neuro protein is not expressed in pancreatic cells and vice versa, even though the genes in a pancreatic cell and a nerve cell are the same.

This is because gene expression is determined early on. Gene expression is what determines the fate of the cells, the function, the differentiation, the growth of the cells, and the DNA synthesis etc, that is why we want to study it. Any change can effect gene expression can effect the cell, can effect certain diseases.

So this is how different cells appear by differentially expressing certain genes. How can we study gene expression? We will learn a couple of techniques.

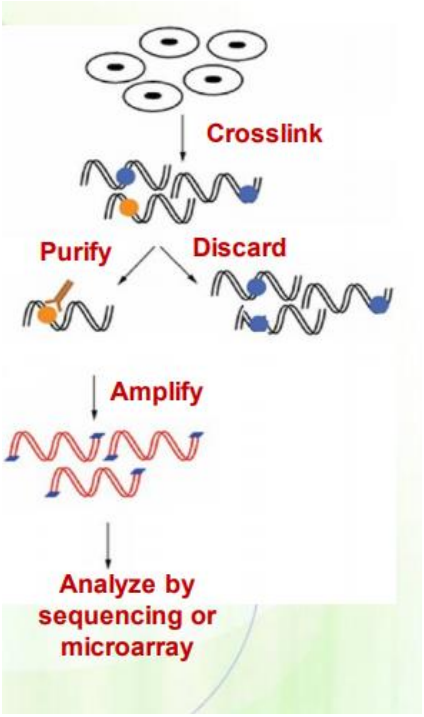
- Isolation of transcription factors using affinity chromatography

The first technique is associated with affinity chromatography, and we learned about affinity chromatography in biochemistry. Basically, it is a column that contains beads, the beads are surfaced by antibodies (the same antibody) and these antibodies are very specific as the interactions between proteins like globulin antibodies with proteins is very specific. We have antibodies that can bind to one single protein very specifically. So what we can do is we take a bunch of proteins from a



cell and pass them through the column, and we will get just on single protein (the one that binds to the antibodies). This is affinity chromatography. But in this case for DNA purposes and transcription factors, on the bead we don't have an antibody, we have a DNA sequence instead, just like the probe. This DNA sequence is specific for a certain promoter, so we add the same DNA sequence to the beads. Now we take the nuclear proteins that include transcription factors and pass them through the column. Only one or two or three certain proteins would bind to that DNA sequence andwith transcription factors that recognize the DNA sequence, anything else would pass unbound. And then we will get the purified protein that we wanted (we eluted the bound proteins). So here we know the DNA sequence and want to find out which proteins bind to it, so we purify it, and study it with the techniques we learned in biochemistry. Again, so with affinity chromatography we want to check which proteins will bind to a certain DNA sequence.

- Chromatin immunoprecipitation



Now if we want to do the opposite, we will use a technique called chromatin immunoprecipitation. Precipitation=ترسيب so what we are doing is that we are precipitating, purifying, isolating, a certain protein using immunoglobulins and that is why it is called immunoprecipitation.

We have a transcription factor, we know it is a transcription factor because we studied the sequence of the gene that makes this transcription factor and saw that it has a zinc finger domain. Zinc finger domain is only present in transcription factors of steroid receptors. Now we want to know to which promoters does this transcription factor bind, if we can know the promoters then we will know the genes that are being regulated by this

transcription factor. Because if we know the DNA sequence that it binds to and we have a data base (the human genome). So if we took the DNA sequence, knew what it is and placed it in the data base. The data base will give us the exact location of the gene, for example chromosome 8 in this location and is right before this gene (it will specify a gene), so maybe this transcription factor regulates the expression of this gene on chromosome 8. If the database show that the gene is present in multiple locations then this means the transcription factor regulates multiple genes.

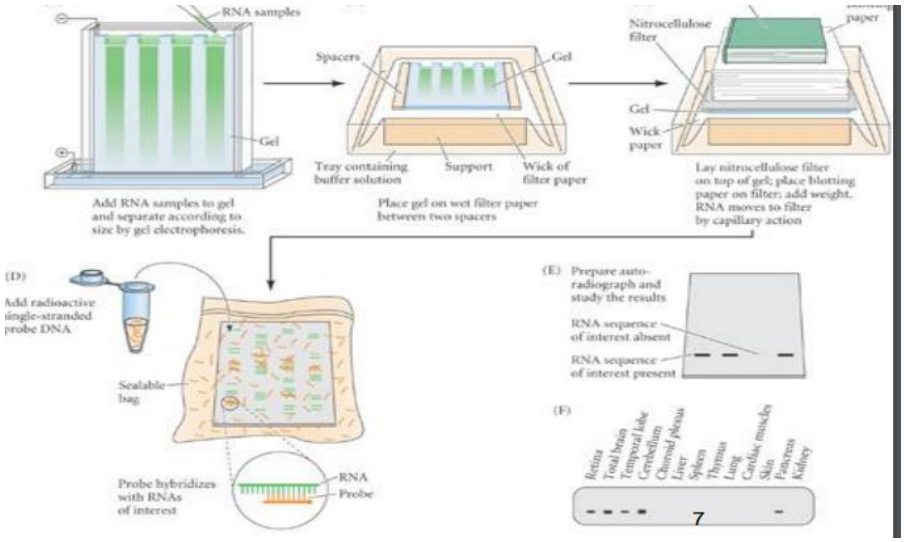
So how can we determine the DNA sequence that this protein binds to?

The idea is that we take the cells and cross link them (chemical linkage between the protein and the DNA) and all of the proteins bound to the DNA becomes cross linked (covalently bound to the DNA). We take the protein that is bound to all of the proteins, all of the unbound proteins are discharged as we only want the proteins that are bound to the DNA. We add enzyme (DNAse) and it breaks down the DNA to smaller fragments. Then we add the antibody, it binds to the transcription factor (that we know) and it pulls it. So, we are separating this particular transcription factor using the antibody, we discard the other DNA sequences that are not attached to transcription factors. Then we amplify (PCR) the wanted gene the one we want to study. First we remove the transcription factor and reverse the covalent linkage and then amplify it. Then we can sequence it or we can do a micro array.

Now we know the DNA sequence (the one that the transcription factor binds to), now we take it and add it to the database and the computer will tell us exactly where it is and we will know exactly which genes can be regulated by this transcription factor. We want to study the gene expression even more, so we use northern blotting.

- Northern blotting

This is a technique created to analyze gene expression,we separate RNA in a gel, transfer them to the membrane, the probe attaches to certain RNA sequence (complementary) and then we have a signal.



WILL BE CONTINUED IN THE NEXT LECTURE