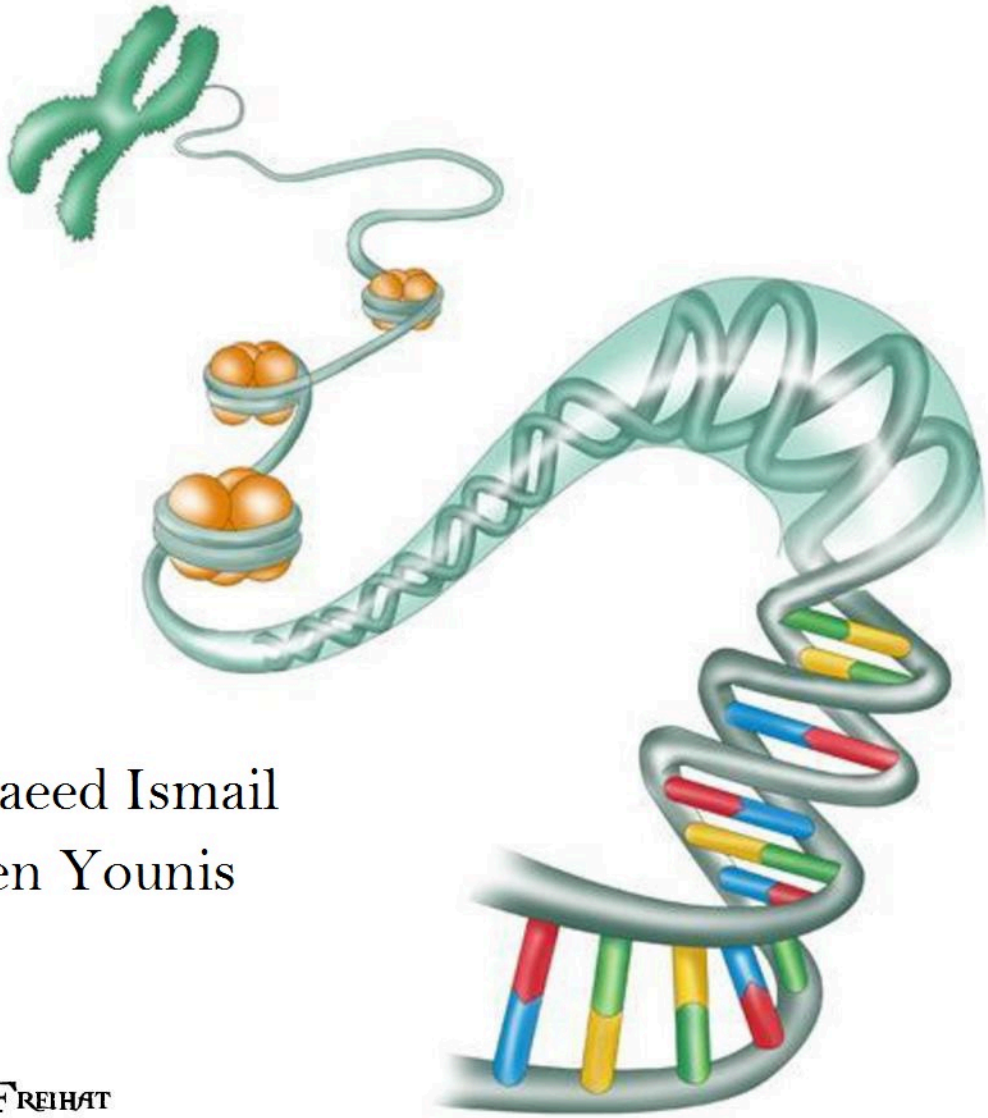
University of Jordan
Faculty of Medicine
Batch 2013-2019

Medical Committee
The University of Jordan

Genetics & Molecular Biology

○ Slides ● Sheet ○ Handout ○ other.....



Sheet#: 20
Dr. Name: Saeed Ismail
Done By: Leen Younis

Designed by Nadeen Al-Freihat

Regulation of Gene Expression

We agreed that in each human cell, we have about 25,000 genes. Cells in our bodies are differentiated; they express only few genes that are needed for the function of that cell.

The *expression profile* is: which genes are expressed in that cell at that given time. The expression profile of a skin cell is different than that for a neural cell, even though they contain the same genome sequence (the same 25,000 genes), they lose control over some genes that they don't need for their function and retain the control over other genes that they need.

Cells are able to switch on/off certain genes when needed according to:

a. Development period: (e.g Fertilized egg-------adult human)

b. Differentiation stage: (e.g Stem cell--------Lymphocyte)

c. Environmental changes (e.g UV and melanin production)

So how do we control these genes that are needed for each cell?

- **The control of genes in prokaryotes differs from eukaryotes in certain ways:**

1. Having several genes under the control of one promoter – Operons -, so they produce one RNA, *Polycistronic mRNA* that can be translated into several proteins that are needed for the same function, which saves time and space.

For example: they could be enzymes in a pathway that utilizes lactose. Normally the bacterial cell uses glucose for energy, in the absence of glucose cells use lactose instead, so they need the right enzymes to get lactose in, break it into glucose and galactose and then convert galactose into glucose.

All these enzymes will be needed under the same conditions so that's why we can put them all under the control of the same promoter.

2. Downstream of the promoter and near it, there's a sequence called the *operator*. Related to the operator, you can find:

- Repressors, they bind to the operator and stop the action of transcription normally by blocking RNA polymerase binding, must be removed when we want to transcribe an operon.

- Inducers, once we want to transcribe a gene, inducers are expressed and they remove the repressors so that the RNA polymerase can do its job.

- In other genes that are used a lot, there's no repression, RNA polymerase can always transcribe, once we need to stop the transcription of the operon, we activate the repressor by a co-repressor.

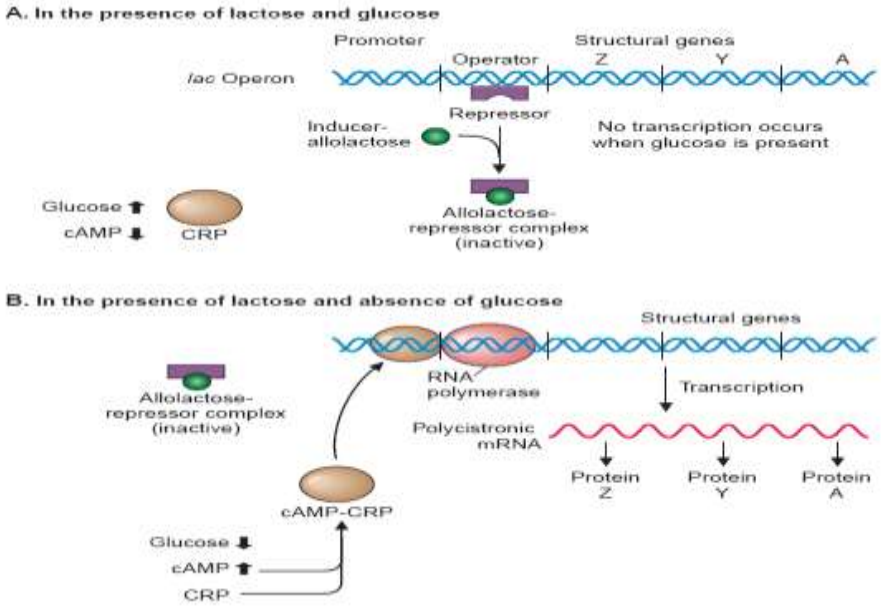
Lac Operon (lactose operon)

It's one of the first operons that were discovered in bacteria, it gives the enzymes needed for the utilization of lactose. Bacteria that uses glucose as a source of energy, even if it had lactose available, that won't be enough to activate the operon. There has to be a decrease in the level of glucose **in addition to** the lactose available.

How does the cell know all that (I don't have glucose & I have lactose, so I can switch on the lac operon to utilize lactose)? And how does it translate that fact to a molecular level?

Some of the secondary metabolic events occurring due to the low glucose level will decide that. If you have low glucose, you're most likely going to have high cAMP and that will react with a protein called the CRP, together (cAMP & CRP) they can bring the RNA polymerase to transcribe the operon.

But that (low glucose level) is not enough, Lactose should also be available since one of the lactose derivatives; allolactose will work as an inducer for lac operon, by binding to the repressor that's usually bound to the lac operon operator, remove it thus inactivate it.



QUICK SUMMERY:

In bacterial cells, we expect the lac operon to be **repressed** since it's using glucose. If there's *no glucose*, there's *high amounts of cAMP* that will bind CRP, together as cAMP –CRP, they bring the RNA polymerase and transcribe the lac operon that will give 3 enzymes to get energy from lactose.

But that's not enough, for RNA polymerase to come; the repressor should be removed by the inducer (allolactose) that's originally from lactose when there's a high amount of it.

One final time :P (⇧ Lactose means the presence of one of its derivatives (allolactose), allolactose works as inducer that takes away the repressor from lac operon operator) + (⇩ Glucose which means ⇧ cAMP, cAMP binds CRP, which brings RNA polymerase to transcribe the operon).

Another example why prokaryotes can be different in their gene regulation:

3. Attenuation (إضعاف) of transcription, weakening of transcription.

Because prokaryotes have no nuclear membrane, there's no reason why translation and transcription shouldn't (the dr. said could, but I think he's mistaken) coexist. While RNA polymerase is still copying the DNA, the RNA is being produced. Resulting in a free 5' end while the 3' end is still being transcribed. Nothing will stop the ribosome from binding the 5' end of the mRNA to start making the protein.

⇨ So the translation and transcription are happening simultaneously.

This can be cleverly used by the cell to control gene expression. For example, another operon called *trp operon* is used to synthesize tryptophan (Trp) which is an amino acid used by some bacteria when there's a low amount of it in the cell.

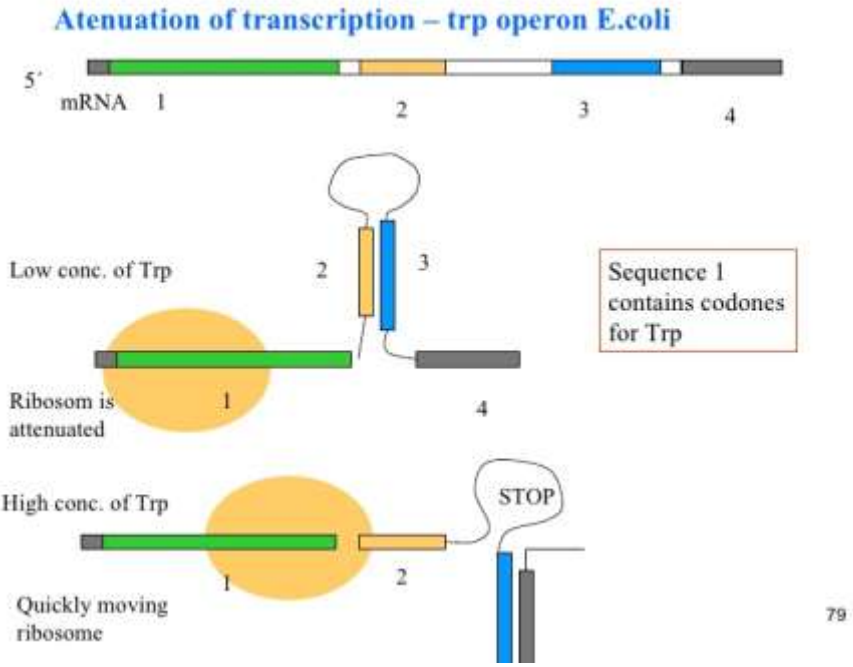
Would the cell synthesize tryptophan if there's enough of it in the cell? No. If it's lacking trp, it would need to synthesize trp and it has to make the enzymes necessary to synthsize trp. These 3 enzymes are put in the same operon because they're needed together. RNA polymerase is transcribing the first codon, the ribosome is waiting for it, it'll read the first codon which needs a tRNA that's carrying tryptophan. If there's enough trp in the cell, it would be easy to obtain it. The ribosome will move to the second codon which would also need trp and if the cell still has enough trp, it's easy to get and so the ribosome will move fast to the next codon and so on.

The first few codons will need a tRNA carrying tryptophan, so if there's plenty tryptophan in the cell, the ribosome will move across the 5' end fast. This fast movement will remove the polymerase stopping the transcription of the trp operon that gives enzymes to make more tryotpohan.

There's no need to copy the enzymes that produce trp since we already have enough of it and that shows from the fast movement of the ribosome.

On the contrary, when there's not enough tryptophan, the RNA polymerase is transcribing, the ribosome wants to translate to protein, the first codon needs trp so the ribosome will stall and stops translating for a while until it gets a tRNA carrying it, the second codon also needs trp so the ribosome waits for another while for more trp and so on. This slow movement of the ribosome will allow the RNA polymerase to complete its job and produce RNA that translates into enzymes that produce tryptophan because there's no enough tryptophan.

The trp operon is a classical example of the attenuation of the transcription. It's arranged in a fantastic way; the first few codons of that operon will code trp which is the target of the RNA. If there's plenty of trp in the cell, the movement of the ribosome will be so quick that it will disrupt the action of the RNA polymerase and it won't transcribe that operon. But if the amount of trp in the cell is low, the movement of the ribosome will be slow and it won't disrupt the RNA polymerase ➔ Transcription will continue, RNA is produced, proteins are made and tryptophan is made.



4. σ Sigma factor (the guide of RNA polymerase)

We talked about it in RNA polymerases of prokaryotes. RNA polymerases are blind and they can't see the transcription factors on promoter sequence, so they need the sigma factor to lead them. Because they need the sigma factor, the RNA polymerases are under its mercy and control. We can control the transcription of those genes if we control the RNA polymerase by hiding the sigma factor. As the transcription factors recognize the promoter and the RNA polymerase recognizes the transcription factors. Whenever we need transcription, we release the sigma factor that guides the RNA polymerase and transcribe the genes. Each group of genes has its own sigma factor.

These are all unique to prokaryotes.

Eukaryotic regulation of gene expression is more complex, because:

1. There's a nucleus so <u>there's segregation between transcription and translation in space and time</u>. We need to wait for transcription to be done to start translating.

2. Our DNA is complex with histones, so to regulate a gene expression, the DNA needs unwinding from the histones and once copying is done, it must be wind again. But in bacteria, its DNA is called the naked DNA as it's loose and it's always ready for transcription.

3. Our cells are diploid so there are 2 copies of each gene. When you want to switch off a gene, you have to switch off both copies. And whenever you think about expressing one, you need to consider switching on either 1 or 2 copies of the gene. While in bacteria it's only one gene.

4. We have 2000 times more DNA than an E.coli. A bacteria has 3 million bases, we have 6 billion bases. A bacterium also has 3,000 genes, while we have 25,000.

5. Bacterial genes have no introns; the coding sequence is all the way from the promoter to the terminator. While in our cells we have to transcribe them to make a very long RNA and then remove the introns and join the exons.

6. We don't have operons, E.coli has 3000 genes to control, many are group regulated, and each 3 or 4 genes are under the control of one promoter. We have 25000 genes, each with its own promoter.

   To make it more clear, if you have1 button for each row of lights in a room, it's easier to switch on the entire room. But if you have a specific button for each light, it will be more complex but more accurate.

   Regulation of eukaryotic gene expression is divided into 5 levels:
   1. DNA Level
      Is this DNA transcriptional and expressible or not, can we express it? If this gene is found in the part of the DNA that's tightly bound to histones, that would make it hard for the transcription machinery to get there and transcribe it. At DNA level, it's not available for expression; we need to unwind it from condensed DNA to diffused DNA and then think about transcribing it.

   2. Transcriptional level, DNA is available and we can transcribe it. It's the most important level, where we decided if we want 10 copies or 10 million copies of the gene; it will be very well calculated. Unless something happens downstream, where conditions change and we get rid of several million copies because we only need a few million.

The transcription control is related to two main elements, and involves the interaction between them:

- The promoter-enhancer regions (cis-acting element)
- Transcription factors (trans-acting element)

3. Post transcriptional level. An example: We made 10 million copies of RNA for those genes then we discovered that we need less than that. Can we change and regulate that? Yes! We can send out to the cytoplasm only 1 million.
   Another example is by not splicing them and keeping the introns there so they won't produce proteins.
   We can also control the alternative splicing.

4. Translational level. The RNA is made, its already in the cytoplasm ready for ribosomes, the cell could decide that there are too many RNA's made and the cell doesn't need that much so it decides to translate only a few copies only.
   Or it decides to make more copies of the protein so instead of one copy making 10 proteins, one copy will make 100 copies of proteins. How? Ribosomes will work faster.

5. Post translational level (protein level).
   RNA was made (transcriptional level), processed and sent out the nucleus (post-transcriptional level), and translated to a polypeptide (translational level). Since most proteins are modified, for example, some are phosphoprotein; they wouldn't work without a phosphate group so we can control them by removing the phosphate group. If the cell discover that I don't need all that proteins → they've been made so what shall I do? Just take the phosphate group out of the protein and thus the protein is not functional. The cell could also discover that it needs more copies so it reattaches the phosphate group back so the protein is now functional.

DNA level

The DNA level includes:

- Chromatin remodeling
- DNA methylation
- Gene rearrangement (very few cells, eg. B-cells)
- Gene amplification (only some cancer cells)
- Gene deletion (only in RBC's)

The first two are the general ones found in **all** cells that I want you to understand and remember. And the other 3 are just exceptional ways by which **some** cells control their gene expression at the DNA level. Gene amplification is an abnormal process.

- **Chromatin remodeling**
  - It is the conversion between condensed (heterochromatin) and diffused/loose chromatin (euchromatin) which is true chromatin.
  - It's reversible, you can wind and unwind the DNA around the histones core hundreds of times.
  - Happens by two ways:
    1. ATP driven, few cells use it since it consumes energy and it's not efficient.
    2. Histon acetylation method, the majority of cells use it.

    First of all, the DNA is attracted to histones since DNA is negatively charged and histones are positively charged since they're rich in basic, positively charged amino acid, lysine.

    ⇨ (negative/positive interaction).

    The DNA detaches from the histones core by masking the positive charge on the lysine side chain by adding an acetyl group using an enzyme called *histone acetylase*, which takes that acetyl group from acetyl coA.

When the positive charge (on the lysine side chain) is masked (by an acetyl group added by histone acetylase enzyme) → no positive charge → the DNA gets loose (euchromatin). To return the DNA back so it attaches to histones, we expose the positive charges by removing the acetyl groups by *histone deacetylase*.

All of our cells can perform this chromatin remodeling and it's reversible. However, in RBC's, it's irreversible. As it's maturing, the RBC will eventually lose the whole DNA content. So in different stages while it's differentiating, it needs less and less genes until it needs only globin genes since it's the only protein it's interested in in its final stages of its life to make hemoglobin. So the RBC's convert the vast of its DNA that it doesn't need any more into heterochromatin (condensed).

⇨ The RBC's will get rid of most of its genes by wrapping the DNA around the histones and deleting its genes. At the end of its life it only needs globulin genes because it's the only one it needs to produce hemoglobin.

- DNA methylation

Used widely by many cells to switch on/off genes at the promoter which controls the gene expression because transcription factors identifies the promoter and these transcription factors decide whether to transcribe that gene and bring the RNA polymerase or not. So it's not the RNA polymerase that would see the promoter, it's the transcription factor that would see and read the promoter and then they bring RNA polymerase.

DNA methylation camouflages (hides) the promoter sequence from transcription factors.

They attach methyl groups to cytosine and some guanine residues since the promoter is very rich in cytosine, if you hide them, the RNA polymerase will hover around the promoter and it won't see them as they are methylated.
Removing the methyl groups will make the promoter visible for transcription factors.

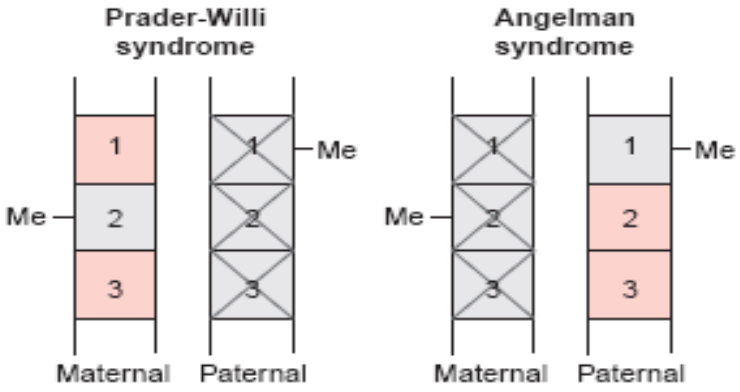
Comparing the RBC with other cells, we'll find that they do methylate their genes as they're differentiating. But they hypomethylate the globin gene.

* Gene imprinting: if you take a skin cell, some of its genes are tightly packed and some are methylated and that's why even thought it has the entire genome, it only expresses certain genes.
This methylation can be used in treating cancer, where the gene we don't want to express is methylated. But it's very difficult due to selectively methylate a particular gene.

**The clinical notes in the book are not required but here's an example:

Males and females methylate their genes differently. That's why if you inherit the same mutation that involves deletion at chromosome #15, it will cause different diseases. If the deletion at chromosome 15 is inherited from the father, that deletion involves genes 1,2,3 so the patient's left with the maternal copy of chromosome 15.
As a mother, the mother has a gene imprinting that is received to her child, females methylate gene #2, so it's not functional, and its counterpart is deleted. So in that case the patient lose both copies of gene #2 (deletion from the father and methylation from the mother for both copies of gene 2) …**Prader-Willi syndrome.**

On the other hand, if he inherits the mutation of chromosome 15 from the maternal side, he's left with that paternal copy of the gene, as a father and because of the imprinting; gene 1 will be methylated, so he will lose both copies of gene 1… **Angelman syndrome.**

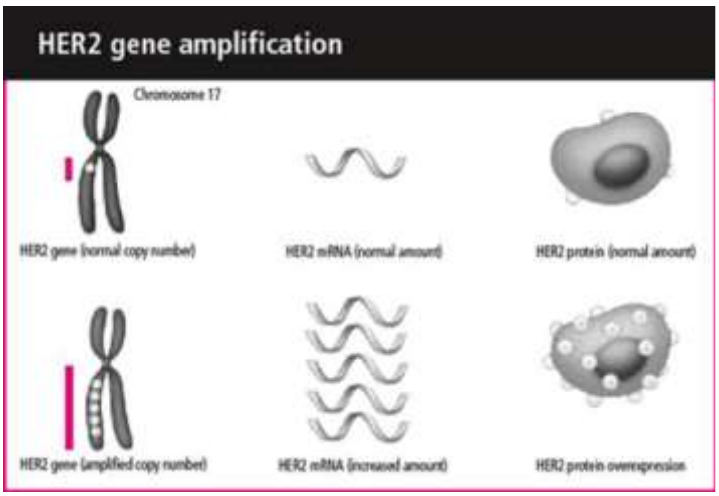


The three other examples affecting DNA level which are not common:

- Gene rearrangements

When we talked about B-cell, at the DNA level, the expression of some genes becomes impossible. This is the story of the mother B-cell which has a huge region at chromosome 14 which enables it to make thousands of different antibodies. When that mother cell decides to pass the ability of making antibodies to a daughter, it won't give the daughter cell the entire genome, it will only pass her a small region. So at the DNA level, the daughter cell can only produce only one antibody, this is a very specific example. Since that mother b-cell has many daughters and each one of them can produce a different antibody, so in total the body can still make all the 10 thousand antibodies. This is good. Since there is dedication! The daughter cell will spend its entire life waiting for the antigen on a specific virus for example to make that antibody.

- Gene amplification
  It's not a normal thing to happen. It's related to some cancers, where they over-express oncogenes which are genes that make the cell divides abnormally. There is a main way to cause the cell overexpress a gene which is to make as <u>much RNA</u> as possible and then make as <u>much protein</u> as possible. But you **can't** affect the DNA level since there are only 2 copies of the oncogene. So cancer cells can make millions of RNA's and millions of proteins but they are limited to only two copies of the gene.
  Some cancer cells and especially in breast cancer, produce many copies of the DNA. They identify the DNA sequence of that oncogene, and make multiple copies.



30% of breast cancer cells identify an oncogene called HER2-NEU on chromosome 17, which codes for a receptor found on normal breast cells. They go to that chromosome that codes for the receptor and make multiple copies of the gene (double minutes) and insert them into the same chromosomes or another chromosomes, they made multiple DNA copies not RNA copies. And this lead to overexpression of the protein which is responsible for making the cell proliferates and divides and this's how cancer cells propagate and then metastasize.

- Gene deletion (exceptional case)
  This is the method that RBC's use at the end of their differentiation. They delete their genes, while normal cells wrap the gene they don't use around histones or they methylate it. This is an exception for RBC's.

THE END.

Written by: Leen Younis