Introduction to Developmental Biology

Prof. Subramaniam K

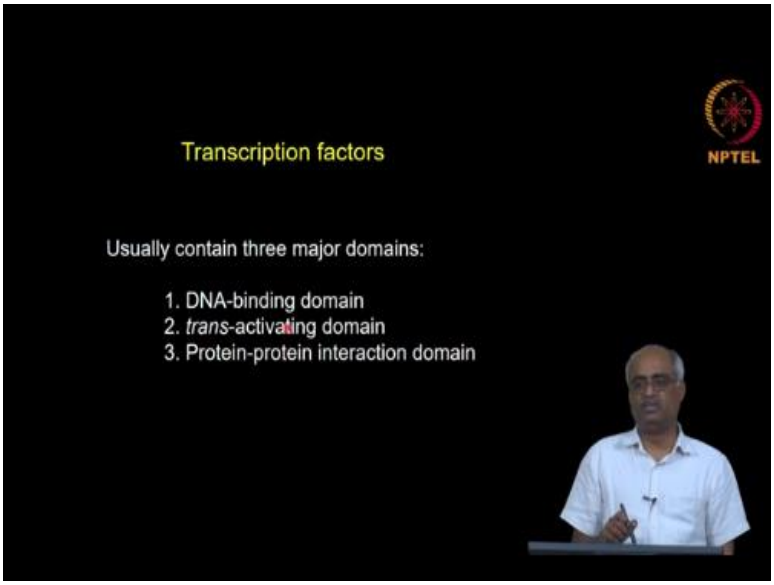
Department of Biotechnology

Indian Institute of Technology, Madras

Lecture No-06

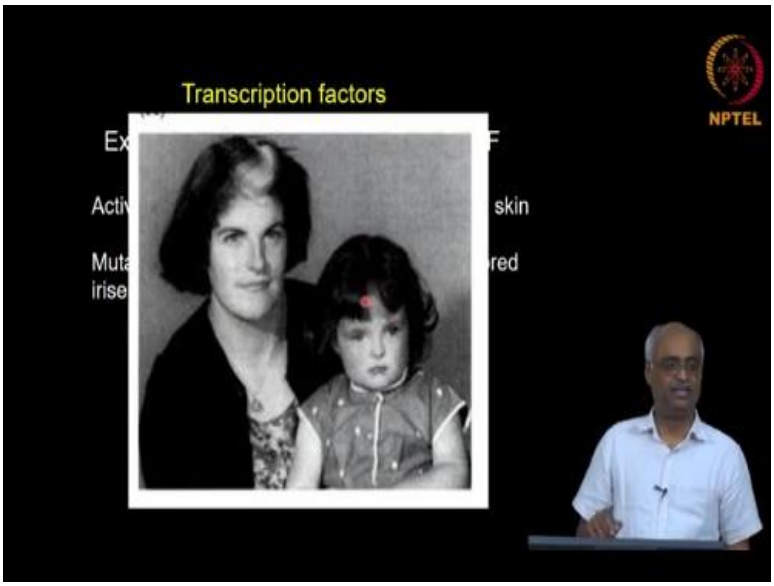
Differential Gene Expression (Part 3 of 4)

(Refer Slide Time: 00:14)



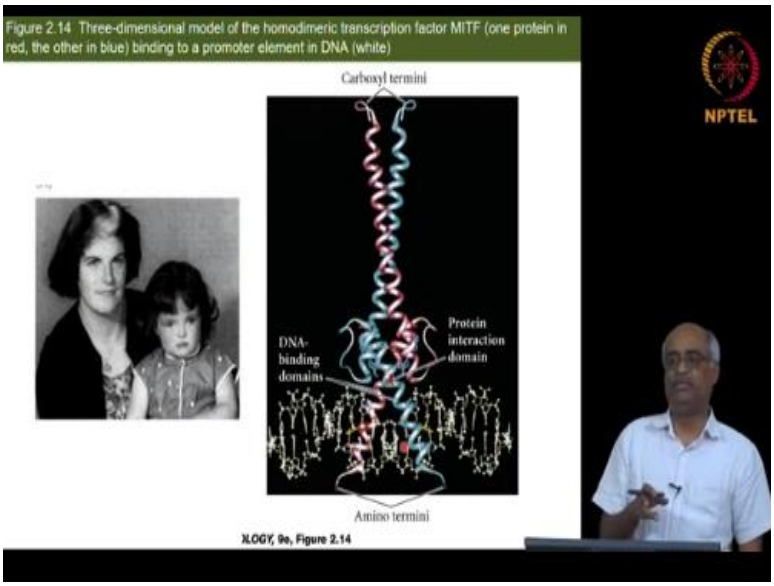
So today, we will focus on transcription factors, and then we will move on to modifications to DNA itself and how they influence differential gene expression and, therefore, development. So transcription factors usually contain roughly three different domains. This does not mean all of them to have all the three all the time. But by and large, most of them have this DNA binding domain, which is a part of the protein. That binds to the DNA directly in amino acids interacting with the DNA double helix and a trans-activating domain where other proteins or factors that bind might modulate the activity of that particular transcription factor and a third one protein-protein interaction. Some of the transcription factors, for example, acts as a dimer, so the two polypeptide chains interact in that region. Sometimes other proteins interact and influence their activity, so the trans-activating ones are responsible for actually activating or not activating the RNA pol II to eventually.

(Refer Slide Time: 01:41)



So any defect in these transcription factors causes disease. An example is MITF, so this transcription factor is expressed in the ear, the skin, and the pigment-forming cells of the eye like irises. If you have a mutation in this, you will have a problem hearing and multicolored irises and then white forelock. As you see in this picture, see the mother has white forelock, and her daughter also has genetically inherited. So, particular problems arise when you have a specific transcription factor missing. So their activity is essential.

(Refer Slide Time: 02:36)

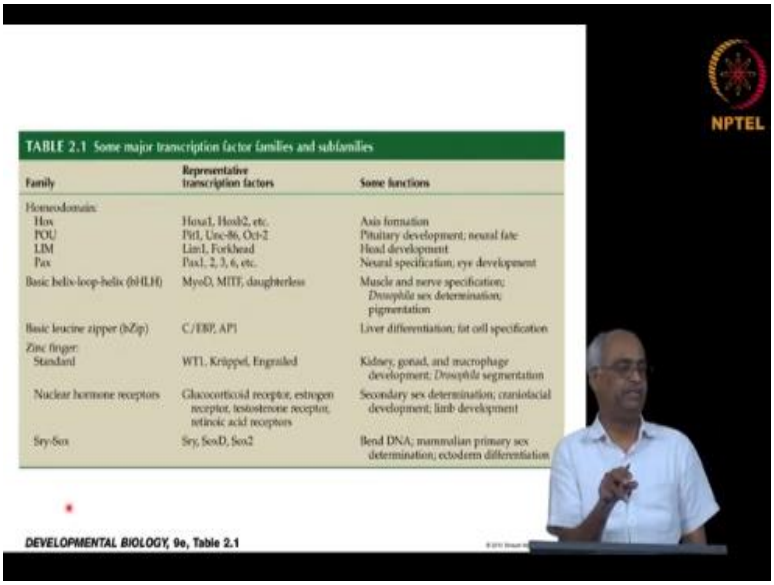


And I am going back to those three domains. So in this model, protein dimerizes in the middle portion, which is the protein-protein interaction domain. So that helps them to dimerize, and the

long carboxy-terminal region is the one that helps in recruiting other proteins. For example, a histone deacetylase, etc., and the amino-terminal region is where the DNA binding domain is located.

So there are different types of DNA binding domains, and based on that, the transcription factors are classified into several classes. Within those classes, small variations in the sequence might define which promoter they bind and do not bind.

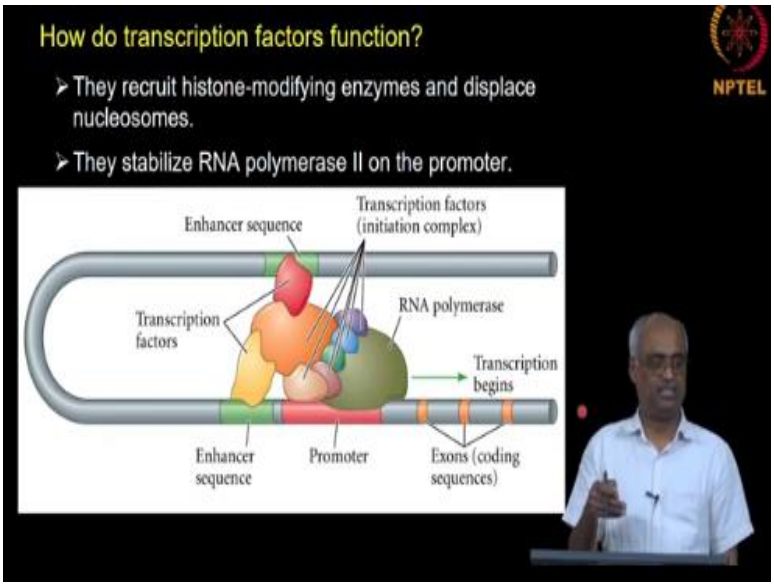
(Refer Slide Time: 03:33)



To give you an idea, let us look at some of them as a table which is there in the book like homeodomain. So that is a particular DNA binding domain that is conserved, and that is present in these proteins listed here. So we will see the Hox protein in detail several lectures later from now, and then some have this helix-loop-helix, HLH, and that is present in these transcription factors, and these are their functions listed here in the rightmost column. And then leucine zipper, they form zipper-like structure based on the leucine present in it. Usually, every 7th amino acid will be a leucine in them, and then you have these zinc finger motifs. These were historically discovered much earlier than others. So this coordinator zinc helps in interacting with the DNA, and this is present in these proteins Krüppel, Engrailed. These are all discovered initially in Drosophila, and the names are based on the mutant phenotype. And they are expressed in these tissues. The nuclear hormone receptors also have zinc finger, and they are

present in the steroid hormone receptors, and then Sry sox that is another domain. So these are the classes based on variations in the DNA binding domain structure.

(Refer Slide Time: 05:06)

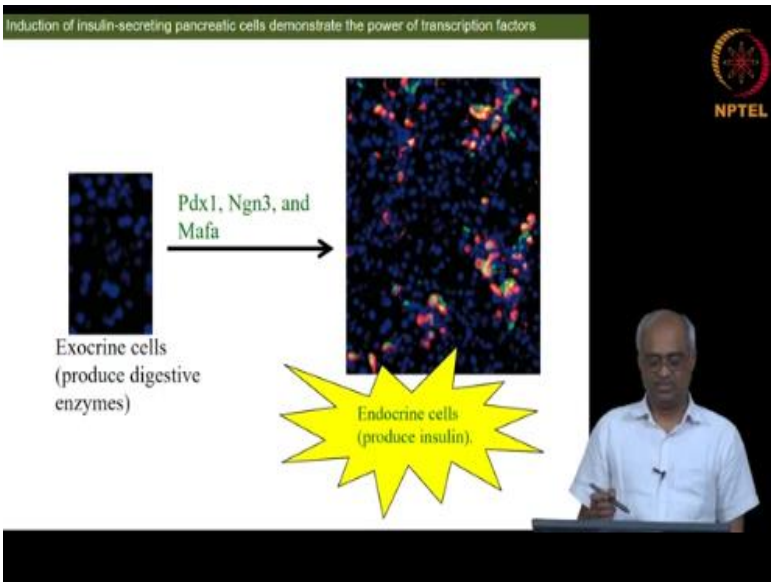


So how do these transcription factors function like how do they activate or inactivate or control transcription? Often it is by one of these two or both, one, they recruit histone-modifying enzymes; for example, when a transcription factor binds to a particular sequence, then this transcription factor may recruit histone acetylase, or they might recruit an enzyme that removes methyl groups inhibiting methyl groups from histones. And by doing that, they will displace the nucleosome structure, and that DNA gets opened up, and it is more accessible for RNA pol II and other transcription factors. So primarily by altering the modifications on histones, they open up the chromatin, which allows transcription, and the second is they stabilize RNA pol II often RNA pol II is bound to the core transcription factors as shown in this cartoon. It is not very stable, but when these transcription factors are bound to enhancer when they interact with all these proteins, they make a more stable initiation complex. They stabilize RNA pol II on the promoter, increasing the probability that RNA pol II will continue to initiate the elongation phase. So in this structure, you see the enhancers can be at a great distance, but through protein-protein interactions, the DNA can loop like this.

So this explains it is present within the coding sequence or in the introns or whatever, or it can be even the downstream sequence. So, these are the general ways, but there are variations for each

transcription factor, but this is generical; if you look at it, these are the primary ways by which transcription factors help in controlling the rate of transcription.

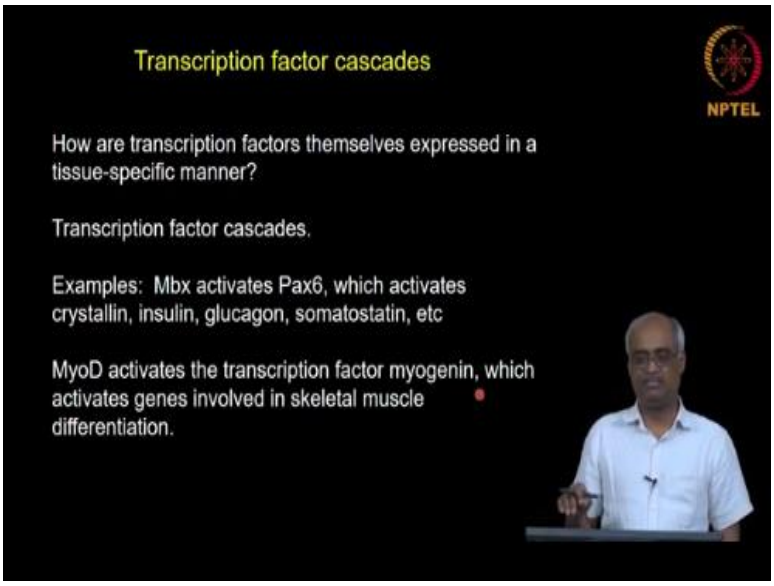
(Refer Slide Time: 07:16)



So how powerful are these transcription factors?. The digestive enzyme-producing part of the pancreas is called the exocrine cells. They usually produce the digestive enzymes, the proteolytic enzymes, and so on, and they do not produce hormones such as insulin or glucagon. So here in the image, this blue is showing the presence of DNA in the nucleus. Now you express three different transcription factors in this Pdx1, so this is expressed in the pancreatic lineage starting from the cells that initially required for the intestinal tube formation. In those cells, if some cells express Pdx1, they set the pancreatic lineage, and in that, if you have these two transcription factors Ngn3 and Mafa, they become the endocrine cells of the pancreas.

Now here you have taken exocrine cells; this is in an organism, it is not in the in-vitro cell culture, so this is in the organism, where in the early on when you express these three transcription factors, you have insulin-producing cells there. So the insulin is stained here with the red color, and one of these transcription factors is fused to GFP, so, therefore, you see green, and wherever both are there, you get yellow. So they are so powerful they can change the fate of a cell from exocrine fate to endocrine fate.

(Refer Slide Time: 09:17)

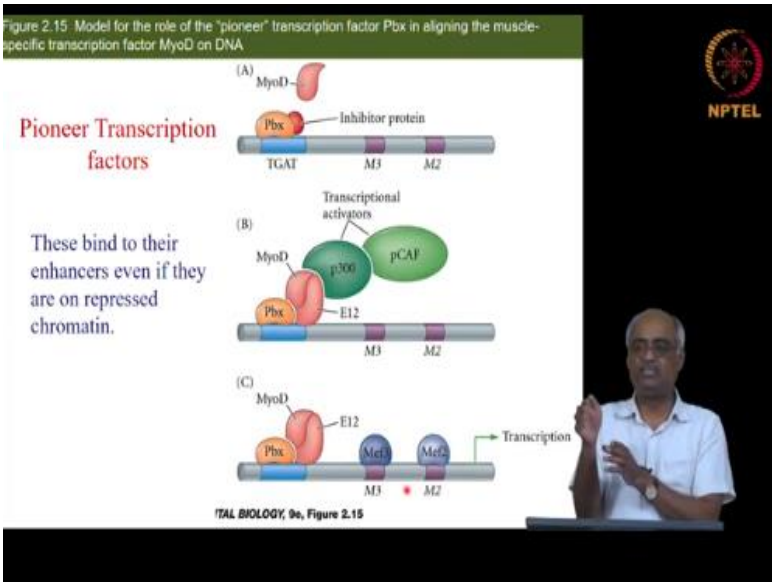


So, of course, now more dramatic things have been done; people have shown by expressing a few transcription factors any differentiated cell can be converted into undifferentiated pluripotent cells. So this leads to a few questions; how do transcription factors themselves get expressed in a tissue-specific manner?

So the answer is quite simple like the stories that people tell, when I was a kid I had one person who was several years older to me like I when I was in elementary school this person was in college, so he talked about some game that he plays. Then I asked who taught you this; he said his PT master then who taught him, then his PT master, so I kept on asking, and I never got the relevant answers because the relevant answer is someone first time discovered it. So similarly, why is this transcription factor expressed in endocrine cells because another transcription factor activated it. Why is that active only in pancreatic lineage it is because another one activated it in the endoderm lineage, So that leads to what is called transcription factor cascades. So they work in the Cascades. Example Mbx activates pax6, pax6 activates crystallin, insulin, glucagon, somatostatin, etc.

Similarly, MyoD, this muscle-specific, really powerful transcription factor activates myogenin, which activates other genes involved in skeletal muscle differentiation. So it is so on and so forth like one after the other. So the central concept here is there is a cascade.

(Refer Slide Time: 11:17)



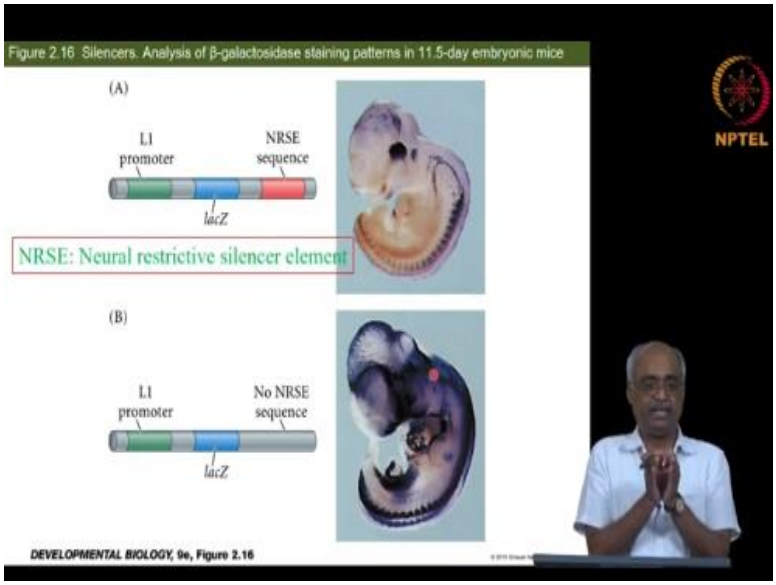
If you follow up the Cascade up to the top, then you have something called Pioneer transcription factors. These transcription factors can open up a highly condensed heterochromatin and initiate transcription. So it need not be already poised for access to proteins. A good example is this Pbx, so it can go and bind to sequences in a highly condensed repressed chromatin.

So that is the definition for pioneer transcription factors. It probably binds to inhibitors bound to that repressed chromatin. But once this transcription factor binds, it can recruit other transcription factors, for example, MyoD transcription factor, and it will come with other accessory factors that help in really activating the transcription finally and open up the place. So that this Mef3, Mef2 etc., can go and bind to their respective enhancers and initiate transcription.

So these are the Pioneer transcription factors, and on top of it you have proteins like the Drosophila Polycomb complex protein and Trithorax. So these proteins bind to the histone modifications and maintain a memory of this original activation, memory meaning when that particular cell fate is specified, and that is going to divide within that individual organism during the ontogenetic stage. All the cell descendants of that particular cell will all know that they have to keep a region active and a region suppressed. So those are done by those proteins, the polycomb, and trithorax group proteins. So this is all about transacting factors controlling transcription.

So like enhancers, there is an opposite phenomenon like there are other DNA sequences that act as negative enhancers meaning their sequence prevents the spread of an activation activity. For example, if an enhancer activates and if it is going to disassemble the nucleosome and spread along the length of the chromosome, then the adjacent genes might also get activated, So you do not want that you want that particular factor to be expressed in that tissue, not all genes. So something has to restrict that activation, and for that, you have DNA sequences to which proteins bind, which insulate or restrict these enhancer activities. So that is what we are going to see next, and they are often called silencers.

(Refer Slide Time: 14:22)

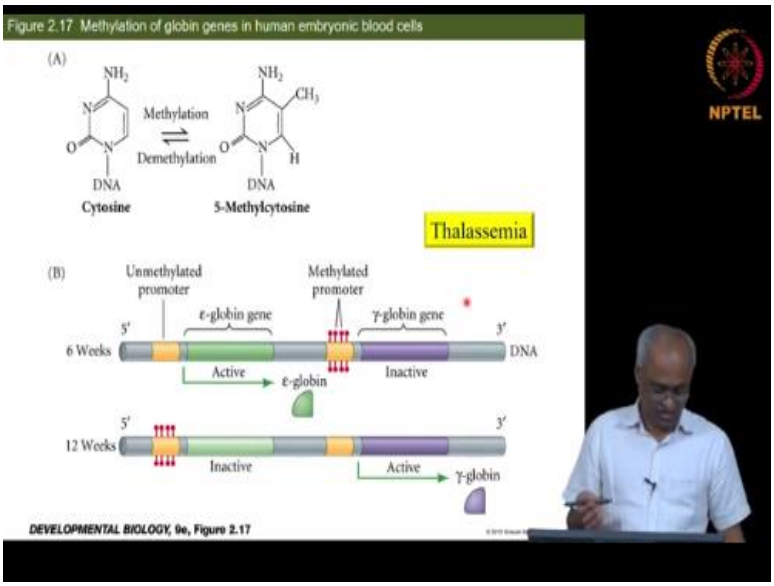


So silencers are opposite of the enhancers, so here is one example, here you have an element called neural restrictive silencer element. So what it does is it binds to proteins, that protein is expressed in all tissues except in neurons. So, as a result, in all the tissues, this sequence will be bound by the protein, and there the genes that are under the influence of this particular enhancer will not be expressed.

So therefore, the genes downstream of those promoters will be expressed only in neurons and as a result this is called neural restrictive silencer. So here in the image is a reporter where instead of the actual gene you have LacZ because you can assay the LacZ encoded protein's activity. So when you have this silencer sequence adjacent to LacZ, you find the reporter is expressed only in the central nervous system here in the 11.5 day old mouse embryo. If you do not have that

silencer element, it is expressed everywhere. So these do the opposite of enhancers; they restrict the influence; otherwise, what will happen is the enhancer effect will not be very specific and restricted to the genes that need to be activated. It will spread and the control will not be really a tight control so adjacent genes may be partially activated etc.

(Refer Slide Time: 16:10)



So next, we go to modifications that happen to the DNA itself. So initially, we saw that methylation, etc. in the histone proteins and that affects the chromatin architecture, whether it is tightly coiled with nucleosomes and histone H1 that is bringing all those nucleosomes together into a solenoid structure or it is going to be opened up for methylation or deacetylation. We also saw some methylation in the H3 tail can be activating. So do not forget that often you may be misled, you will automatically assume methylation means inactivation and acetylation means activation. Acetylation is activation, but that generalization is not for methylation. So now, we are going to look at methylations that happen to DNA.

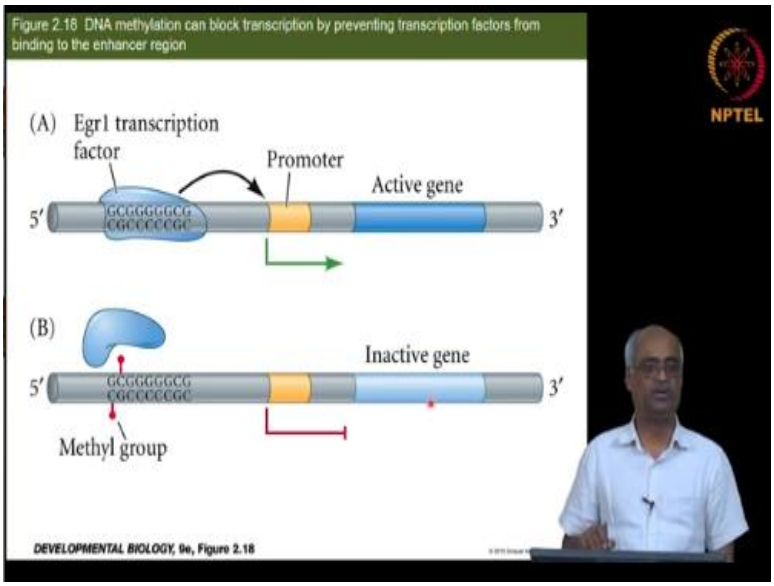
So as mentioned earlier, to perpetuate an active state or repressed state, we have those Trithorax and Polycomb proteins that bind to the modified histones. For example, if something is acetylated and you want that to be active, these Trithorax proteins bind there, and they maintain the active state. Still, very similar but more robust is the modifications that happen to the DNA, and that happens by methylating the cytosine residues. So CH3 is added to the fifth base that is 5-methylcytosine. So this matters a lot in regulation. So here methylation usually means a

repressed state like an inactive gene, and it is not going to be transcribed, and this can be perpetuated through mitotic cell divisions. We will see how that happens in a couple of slides. Second, this can have a developmental time factor involved in it.

Modification happens at a different space and time, not all the time. So a good example of that is the hemoglobin genes. These genes are expressed as ß-globin in the adult. In the early embryo, you have an ε version of the globin gene expressed. Its promoter is not methylated, whereas the γ-globin, which is usually expressed in the fetus, is methylated, so it is not expressed. As the embryo progresses, the γ-globin gets demethylated and gets turned on, which was dormant, and while the ε-globin gene gets turned off and when the infant starts to grow γ-globin gets methylated and is inactivated. In contrast, the ß-globin gene gets activated, and that is what is expressed in our body.

So our genome has ε and γ sequence, but they are methylated and not expressed. They were expressed sequentially during your embryonic and childhood development. So now you have only ß-globin, and there are consequences if there is a problem with this regulation. You may have heard this decease Thalassemia that results from a failure in the sequential methylation and demethylation. So in these patients, you may have a problem activating the ß-globin. Let us say you have a mutation in ß-globin, and now you do not have a functional globin protein produced. Although perfectly good copies of the gene are present in the chromosome, unfortunately, they are methylated. So the gene is not expressed, and that is ß-thalassemia when ß-globin gene that is involved. So this is a very well characterized congenital disease in India, particularly in some pockets bordering Andhra Pradesh and Tamil Nadu. In that area in certain communities where you have marriage among close relatives like first cousin marriage or sometimes an uncle marrying a niece. Like brother marrying the elder sister's daughter. Those are not uncommon; maybe they are rare now, but a couple of generations ago they were not unusual in those families, for example, this sister may be heterozygous, and this guy also may be heterozygous, because they come from the same parent and they survived because they are heterozygous. Now there is one-fourth chance that their child will be homozygous for the mutant allele. So that is how you have ß-thalassemia running in families, and the underlying cause is this methylation issues.

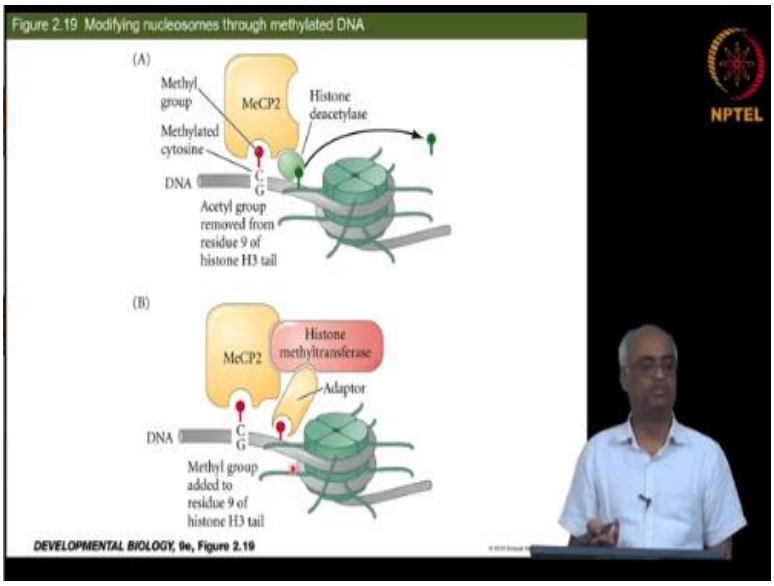
(Refer Slide Time: 22:06)



So now, how do you perpetuate this? So usually, these methylations block transcription by preventing transcription factors binding to the enhancer. Sometimes inhibitors are also involved; they will bind to the unmethylated one, and they will not bind methylated.

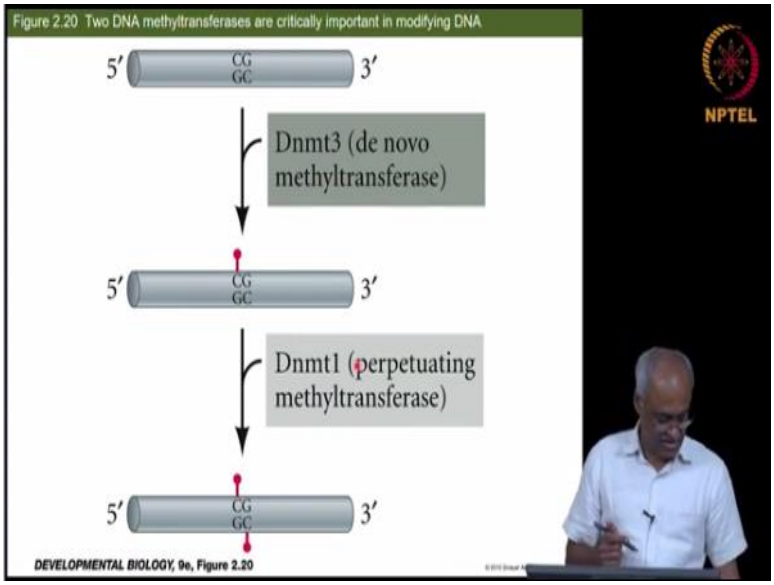
The sequence in this particular cartoon, you have CG coming together. So this is often called CPG Islands, and its significance will become clear in a couple of slides. So for now, do not worry; you think that this promoter region is usually subject to methylation and demethylation. So when it is not methylated, the transcription factor binds and activates transcription from the downstream promoter, and if it is methylated, this transcription factor does not bind; as a result the gene is not active. So, therefore, here, this example shows that DNA methylation blocks transcription factor binding to an enhancer.

(Refer Slide Time: 23:21)



And another way by which they function is, this methylated cytosine may recruit a protein like in this case MeCP2; which can do two things, one removes the acetylation mark by recruiting a histone deacetylase and second recruit a histone methyltransferase and mark histones with inhibitory methyl groups. Due to these two actions, these methylated promoters end up blocking transcription.

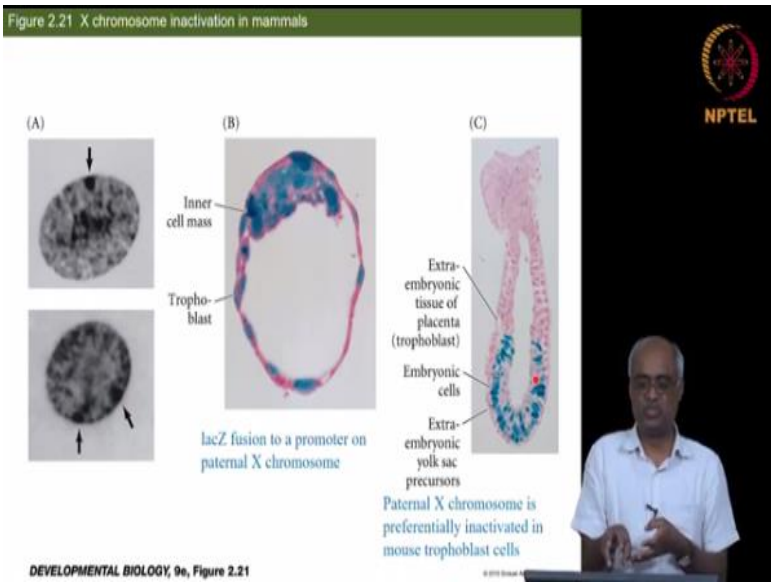
(Refer Slide Time: 23:59)



This sort of methylation based; transcriptional repression can be perpetuated through mitosis. Because these cytosines are always adjacent to a guanosine residue, CPG Island; the phosphate in between probably helps in pronouncing better; otherwise, I would say CG. So normally people call CPG repeats, CPG Islands means in the chromosome here, and there you have a lot of repeats

of CPG. And these are recognized by a methyltransferase called Dnmt3; this does not need either one of the two C's that you see here. CG means the opposite strand will be GC. So you have C in both the strands due to this base complementarity. So here, neither C's are methylated, and this methyltransferase3 can recognize such sequences and that is why it is called de novo methyltransferase. It can methylate with prior information. Now you have a perpetuating methyltransferase. Remember, this methyl group is not erased during mitosis; it is going to remain there. Now, after replication, one strand will have the cytosine methylated the other one will not have. The methyltransferase1 recognizes such methylated cytosines, and they methylate in the opposite strand the nearest C. That is how the adjacent G becomes crucial for this. So now both strands are methylated and again undergo DNA replication, then one strand will be methylated by Dnmt1; the other strand will not be, so this is how the repressed state maintained during cell divisions. So during embryonic development at some point, inactivation by methylation takes place. Let us say the transcription factor cascade and chromosome modification ended up methylating the DNA, now all the cells descending from that original cell will all maintain that active or inactive state.

(Refer Slide Time: 26:30)



And so this has a lot of significant consequences in many situations, particularly here if we look at this dosage compensation. So what is dosage compensation? For example, in mammals like humans, females have two X chromosomes males have only one X chromosome. While the Y chromosome does not have many essential genes, the X chromosome has a lot of important

genes. So will the females produce proteins double the number of males, and will that not cause a problem in terms of the phenotype? So that has to be taken care of, and that happens by one of three different mechanisms. Like for example, if you take *C. elegans,* both the X chromosomes get reduced by half, and therefore you have the final quantities like one, compared to the males; females will have only one X chromosome.

In Drosophila, the single male X chromosome is doubled up. Its chromatin is modified such that it is truly euchromatin, and the output is more efficient. And in humans we do the opposite; one of the two X chromosomes in the female is converted into heterochromatin and repressed. And in this human-derived cell were this arrow points to a large black region is the condensed inactive X chromosome. And this is from a person with three X chromosomes, and therefore you see two black things which are called Barr bodies.

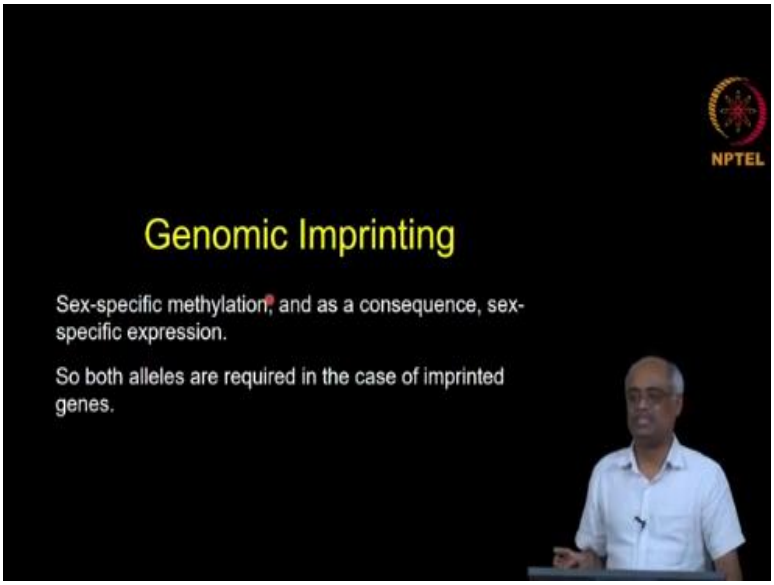
So that is how the inactivation works. The important thing here is if you look at the B and C, in B what do you have is a very early embryo, in this you have the reporter LacZ fused to the promoter on the paternal X chromosome. So LacZ will be expressed if the paternal X chromosome is active; otherwise, there will be no LacZ, and therefore this blue color will not happen. So the pink cells are, where the paternal X chromosome is not expressing. It is not working, so this is very early; you see most of the cells having this color, so this is the inner cell mass from which the entire embryo develops, but when you go to the later stage here in C, these cells do not have the LacZ expression. Later it turns out that in Mouse, the trophoblast cells preferentially inactivate the X chromosome of paternal origin, but in other regions, both kinds mixed. Three essential points to remember about this inactivation is one this starts early in the early embryo meaning in the one-cell stage itself. If it is inactive, then you have an entire tissue or a part of a tissue-derived from this cell having no gene expression. If it is paternal, then paternal expression will be absent. If it is maternal, the maternal expression is inactivated.
Second, the X chromosome gets inactivated randomly, either maternal or paternal. Third, once inactivated, it is irreversible. It remains in the descendants of that lineage, and due to that, you can have patches of variations in the somatic body. And that is often readily visible in organisms where you have a skin color having patches which are seen in calico cats. So these three points

that it happens very early, and it can happen randomly, and once happens, it is irreversible needs to be kept in mind.

All the descendants will have the inactivation, and if this is the case, then if I have a gene on X chromosome and that is very vital. If that gets inactivated in my father's genome, that means having one wild type copy from my mother is not going to be enough. For certain genes, the mother's copy is required. And similarly, for some genes father's copy is essential.

So this is where you will find that; when you are drawing punnet square, it does not matter where the allele comes from either maternal or paternal. But there are situations like this X chromosome dosage compensation where that matters. So we will see that in the next set of slides.

(Refer Slide Time: 32:36)

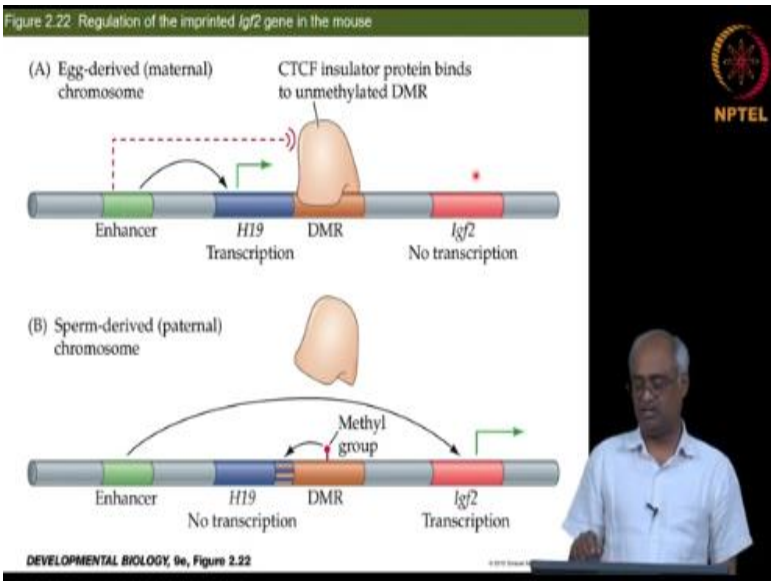


So existing methylation gets erased during gametogenesis, and new methylations take place, and this does not happen to all genes. There are specific genes that are methylated depending upon whether it is in a male or female body. For example, some genes may be methylated during spermatogenesis only, and some other genes may be methylated only during oogenesis. Usually, they are mutually exclusive genes. The genes that are methylated during oogenesis are not methylated during spermatogenesis and vice versa. This is called genome imprinting, so genome imprinting means; sex-specific methylation as a consequence, sex-specific expression.

To further explain, If a particular gene is methylated during spermatogenesis, then it will not be expressed from the paternal allele in the offspring. If it is methylated during oogenesis, that specific gene is not going to be expressed from the maternal allele. So if you have two alleles, assume both are wild-type, one allele should be inactive, and that takes care of the dosage compensation. For example, If a maternal allele is inactive, it does not matter even if its wild-type the paternal allele is required for the normal function of that gene, and the same logic holds in the opposite direction too.

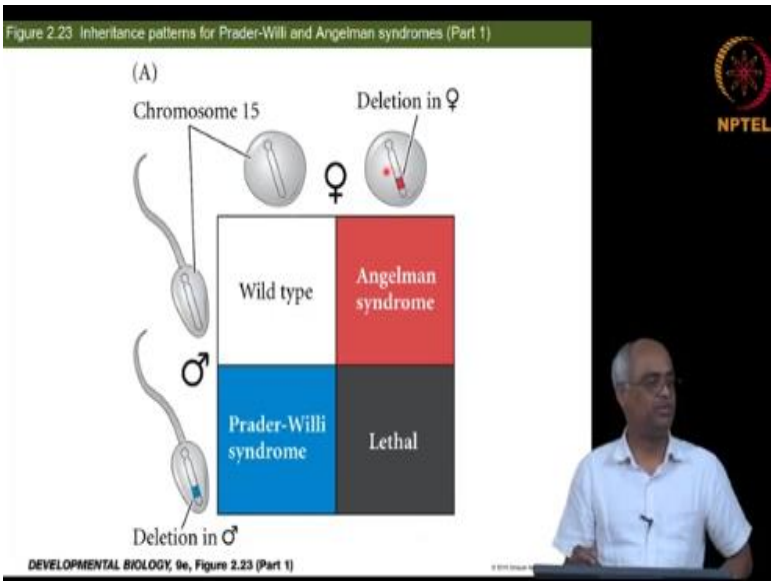
But germline uses a very different set of combinations of the existing molecular biology rules in taking care of its genome. You cannot readily mess with its genome, and the way it protects has its peculiarities, and one of them is this erasing all the methylation and then newly methylate the genes. The new methylation is going to be based on whether you are a female or male. For example, a particular base pair sequence could have come from your mother; now, if you are a male, you will have male-specific methylation added to that gene, even though it initially came from a maternal or female source and vice versa. So the methylation is sex-specific, sex of the individual in whom the gamete is forming. Now, as a result, the imprinted genes, meaning sex-specific methylated genes, need to have both the alleles because one allele is going to be inactivated, and the other allele is going to be the only one available. So, in that case, if you have a mutation in such genes, then you will have a sex-specific phenotypic outcome. So that is what we are going to see next.

(Refer Slide Time: 36:16)



So in this cartoon, in (A) Igf-2 is not transcribed because an insulator protein binds to the Igf-2 sequence and prevents its expression. So this protein binds to unmethylated DNA. Remember in the previous case, like several slides ago, we saw a situation where the activator binds to unmethylated DNA and not to methylated DNA. So here, methylation inactivates the transcription. So, this Igf-2 will not be produced when this sequence is not methylated in the maternal genome. Igf-2 is methylated in the paternal chromosome, meaning this locus gets methylated during spermatogenesis and not during oogenesis. Now an insulator protein binds to this sequence; these are the proteins that bind to silencers, which binds and insulates this coding sequence from the effect of the enhancer. So this just happens to share the same enhancer, so we need not worry here. In the Sperm-derived chromosome, a methyl group is present; as a result, this insulator does not bind, and the enhancer activity impacts transcription from Igf-2, and it gets expressed. So you need to have both copies of it. If you had this mutated like the father carried the mutant version, then due to mutation, you will not have Igf-2 protein, and although your maternal copy is wild-type, it will not express. So due to that, you are going to be deficient in this protein.

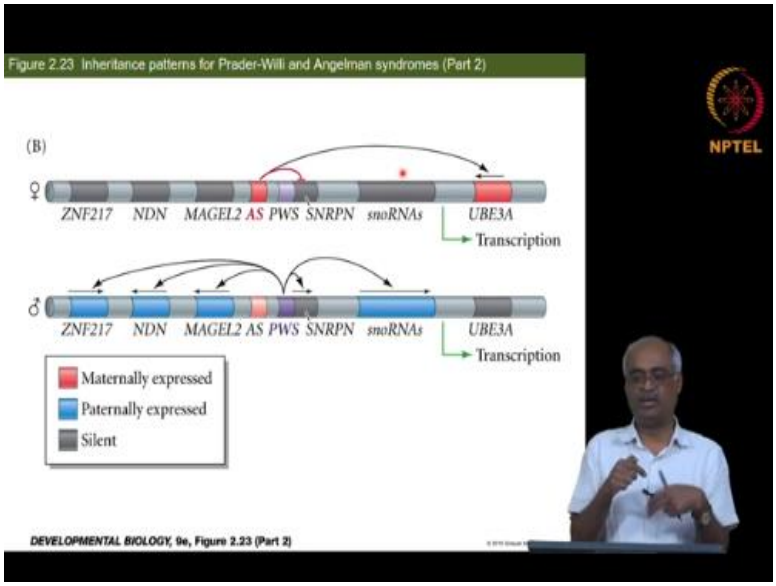
(Refer Slide Time: 38:25)



Such mutations are not hypothetical; it does have real-life consequences. So if you take these two syndromes, Angelman syndrome and Prader-Willi syndrome; these have two different phenotypes coming from mental retardation and seizure and so on. Though both have defects, Angelman syndrome is more severe than the Prader-Willi, but double is Lethal.

So if you look at the first one in this Punnett square you have chromosome 15. This is coming, let us say from a wild-type sperm and a wild-type egg, then you are totally wild-type. Now let us say you have a particular region in 15 that is deleted in the male. So, now you have a wild-type copy of chromosome 15 coming from the egg but that is not helpful. Because the required genes are actually inactivated due to maternal specific imprinting and when you do not have that from the sperm then you get this disease. It is because of those particular genes, usually expressed from the paternal copy.

Now if you look at the opposite; where you have wild-type chromosome coming from the male but a deletion coming from the female, a different set of genes are affected because the methylation is sex-specific here. And now the corresponding alleles in the male are inactivated and you need it from the maternal copy and that is not available due to the mutation and due to the difference in the genes affected you get Angelman syndrome. So this locus is shown in the next cartoon. **(Refer Slide Time: 40:30)**



So this is a complex locus where you have several genes there. The grey indicates inactivation except for PWS in males, this is a typo or a drawing error they had in the books. So PWS is active in males, due to methylation, PWS activates other genes in the paternal copy. But in the maternal copy due to methylation, you have a block in PWS, and because of that, these genes are not expressed. So in this particular case, UBE3A is expressed only from the maternal copy while these blue ones are expressed from the paternal. When you need both the gene products, then you need to have both maternal and paternal alleles, and that is why you get those diseases. So this is genome imprinting which is a consequence of sex-specific methylation.

The critical thing you need to remember is existing methylation marks are erased during gametogenesis. So right now, in your cells you will have imprinting, the paternal allele will know that it is a paternal allele because of a paternal specific methylation. Similarly, the maternal allele will know that it is maternal because of maternal specific methylation or the lack of in both cases, and when both are there, it is fine.

Now when germ cells enter into gametogenesis, these marks are erased, and no methylation happens. So if it is spermatogenesis you are going to methylate a specific, loci, or if it is oogenesis, then you are going to imprint again another locus; these are mutually exclusive. So I am stopping here if you have questions feel free.