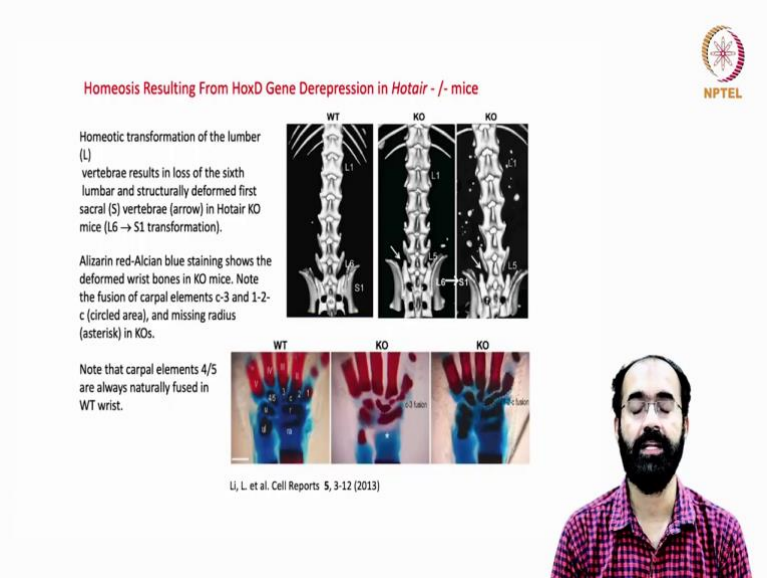
RNA Biology
Prof. Rajesh Ramachandran
Department of Biological Sciences
Indian Institute of Science Education and Research, Mohali

Lecture - 49
Dosage Compensation, Xist and ncRNA in Imprinting: Different ncRNAs and their Roles

(Refer Slide Time: 00:22)



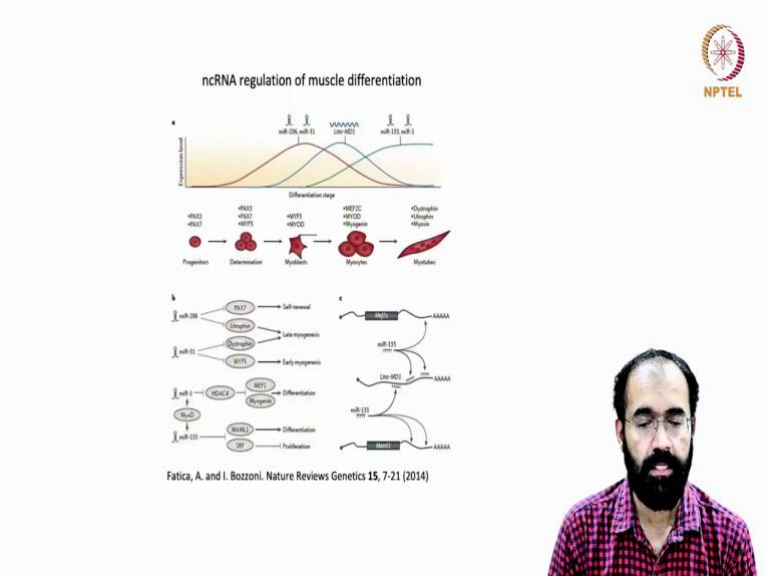
Hello everyone, welcome back to another session of RNA Biology. So, we were here in the previous class that the de-repression of HoxD leads to Homeosis. And the de-repression is happening mainly because of absence of the non-coding RNA hot air. What is homeotic transformation? The change in the structure with another structure, a given place a structure should be there and that is missing instead another structure is appearing there.

And it can appear in the form of a completely transformed new structure like you saw in the case of fusion bond, fused bonds are missing normally in the wild type as you can see here. They are new structure whereas, here the homeotic transformation you are seeing at the place of L5 you have got S1.

So, sometimes L5 is replaced by L6 itself; so, you have two copies of L6. Sometimes the place of L6 you have sacral 1; so, L6 is missing, lumbar 6 is missing. So, different types

of homeotic transformation are possible based on which tissue conditionally you are abolishing the hot air gene.

(Refer Slide Time: 01:51)



So, let us see another situation in which the non-coding RNA regulate the muscle differentiation. So, muscle differentiation is something interesting in the sense that muscle are formed as single cells and later they fuse the cytoplasm will fuse. And they retain the nucleus and we call muscle fibres as syncytial; syncytial means, one cell have got multiple nucleus.

Actually, it was multiple cell to start with and it fused to make a long cell elongated cell and it is the structure of a muscle fibre. And later on, you know more of the cytoskeletal elements will acquire the contractile properties like troponin, tropomyosin actin those kind of myofilaments come into picture for the contraction. And you can read about the sliding filament theory of muscle contraction for more details.

But we are talking about the formation of the muscle muscle differentiation. So, to do this there are different kinds of non-coding RNA comes into picture such as miR 206, miR stands for micro RNA and miR 31 and LNC MD1 and miR 133 and miR 1. So, you can see here this is the expression level in the y-axis, the rate of expression of a given non-coding RNA.

And this is different stages of differentiation starting from the beginning till the end. So, you can see some of them such as miR 206, miR 31 are expressed and it forms a parabolic graph peaks. And then comes down and some of them LNC MD1 they are little late onset and the peak also late and then it comes down.

And for some other micro RNA miR 133 and miR 1 comes much much later. In fact, they are induced at a time when the miR 206, miR 31 is in its peak. That is the time the other microRNA start their expression and they continue to be present; they do not have a decline phase they continue to be present in the differentiation.

So, now we know the muscle gets its identity because of the expression of various muscle specific genes. So, now let us see the progenitors of muscle and then they have to expand in number, progenitors means they are somewhat stem cell like characteristics. So, they are progenitors they can give rise to see just like if you can say a person eventually becoming a doctor, but in childhood like a baby he look the same and same way he saw on another neighbour's baby he will become a advocate.

So, one became doctor another became advocate; so, but at childhood they look the same; so, we can technically call this childhood phase as the progenitor phase. So, this doctor could have also become advocate or the advocate also could have become doctor if tuned in the right way so, at a stage where they have no identity that phase is called progenitor, but eventually they will become something.

So, here also we are referring to as progenitor and progenitors have got some unique set of gene expression that is called PAX -3, PAX -7 these are some of the marker genes that is expressed in the progenitors. And later on, you start getting the expression of PAX -3, PAX -7 and MYF -5 which is a muscle specific transcription factor. And keep in mind this PAX -6, PAX -7 these are all transcription factor paired homeobox gene basically PAX stands for.

Then comes the MYF -5 and myoD at myoblast stage; so, when the differentiation begins then they reach a stage called myoblast. And then the myoblast mature into myocytes and at myocytes you have expression of MEF -2C, myoD, myogenin etcetera, these are also transcription factors. And finally, they will become myotubes which give rise to the muscle fibre.

So, in this stage you have got the expression of Dystrophin, Utrophin , Myosin etcetera. So, you have some unique housekeeping genes expressed at various stages of the muscle development. Now, comes the question how are the different non-coding RNA comes into picture.

Like you can see here different stages of development corresponds to this like individual group of cells are there this is the final maturation. So, you can see this is the beginning stage, in this graph this is the early phase, then this is mid phase, this is the late phase, this is the end phase.

So, each of the non-coding RNA that are expressed influence the expression of one or the other set of muscle specific transcription factors and these are muscle specific structural proteins. So, let us see the miR 206 it is capable of repressing the PAX 7, you know PAX 7 is expressed here. So, the moment miR 206 turns on as you see here it will repress the PAX 7 and also repress the utrophin gene.

So, this repression of normally the PAX 7 is important in the self renewal. So, the repression allows differentiation and utrophin and dystrophin these are all important in the late myogenesis. So, you do not want the late myogenesis gene expressive right now. So, they have to be expressed at a time like a student going to kindergarten UKG or first grade do not want to study any MBBS topic or any you know LLB topic because that has to be taught when they reach that class or that education.

Same way you do not want any muscle specific or the identity genes, but they have the potential to express because this is my muscle progenitor; so, you have to repress that also. So, it is not only repressing those genes that is allowing the multiplication of this progenitor, but also preventing the expression of later expressing genes such as the utrophin, dystrophin, etcetera. So, miR 31 also cause the repression of dystrophin and also repress the MYF 5 and MYF 5 is early myogenesis.

So, you do not want those events also to happen and what miR 1 does? miR 1 inhibits the HDAC 4, HDAC 4 is a Histone Deacetylase, Histone Deacetylases causes the compaction of the heterochromatin or the cause the production of heterochromatin compaction of the chromatin. So, this in turn the HDAC 4, which in turn can influence inhibit the MYF 2 and myogenin expression and because they are important for the differentiation.

And the MyoD is the causative factor transcription factor for the expression of miR 1 as well as miR 133, because this is a muscle specific transcription factor; so, miR 133 what it will do? It will inhibit the MAML1 and SRF1. So, these are MAML1 are important in the differentiation of the muscle fibre and also SRF1 is important it is inhibitor of cellular proliferation. Because, if proliferation is continuing differentiation will be affected; so, you want differentiation to happen.

So, the different micro RNAs can influence various proteins and transcription factors that can allow a unwanted or that can facilitate an unwanted performance in a given stage. Just like a kindergarten student should not be studying MBBS topic, same an MBBS student need not be or should not be studying kindergarten content also. So, same logic applies when you are talking about the differentiation.

When a differentiation is occurring proliferation is unwanted and you do not want the expression of those genes that is helpful in the proliferation of the progenitors. And you also do not want the late expressing genes now; late expressing genes will be expressed later. You do not want right now.

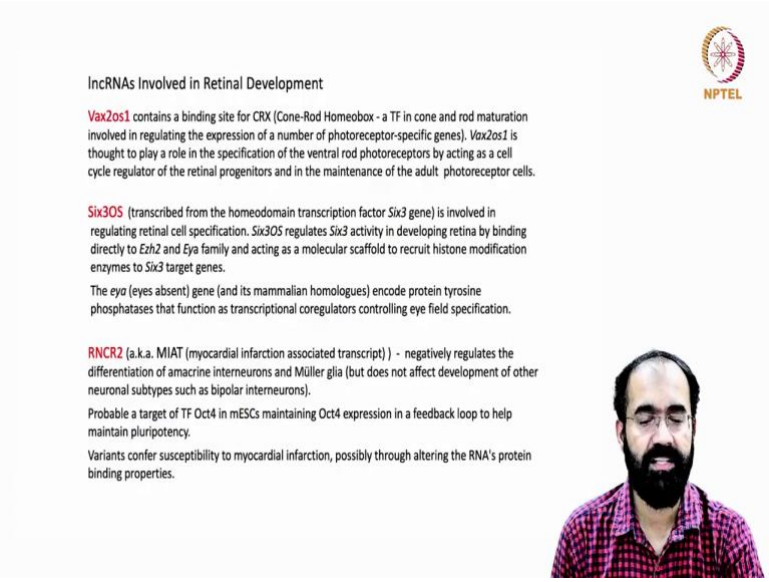
Just like you are going for a party or a dinner you do not want to have desert first, desert should be after the main course food right. Same logic; so, you have to say ok I am done with the starter and I do not want desert right now; let me have the main course. Once the main course food is done then you go for the desert items, same logic has to be followed in the differentiation step also.

So, what this mainly the non coding RNA when they are expressed, they can influence the expression of like here you are seeing MEF 2 C and this is a non coding RNA miR 155 and then Linc, MD1 and also miR 133. So, they can influence the expression pattern of a lot of muscle specific genes. So, in general what we should understand from this studies example more than the non coding RNA and the mRNA pair.

The identity of a given structure whether it is a bone or whether it is a formation of a differentiated muscle in a given location that is regulated and directed by a specific set of micro RNA or non coding RNA which in turn are turned on by that particular tissue specific transcription factor.

Like, you saw here myoD's expressing causing the expression of the micro RNA miR 1 and miR 133. So, what we should understand is the expression of a given set of genes in the specific window of time not too early not too late. That is what the hallmark of a normal differentiation or a smooth differentiation for any cell type; so, here we are using muscle as an example.

(Refer Slide Time: 12:59)



So, now let us see the lnc RNAs involved in retinal development. Retina is the film that is a part of your eye where you have the where you have your images are formed whatever you are seeing the image is formed in the retina. So, what we have to see here is one example is Vax 2 os.

So, this contains a binding site for CRX, cone rod homeobox it is a transcription factor Tf stands for transcription factor in cone and rod maturation that is involved in regulating the expression of a number of photoreceptor specific genes. And Vax 2 OS1 is thought to play a role in the specification of ventral rod photoreceptors by acting as a cell cycled regulator of the retinal progenitors and the maintenance of adult photoreceptor cells.

So, this non-coding RNA is making sure that Vax 2 OS 1 non-coding RNA is making sure that the ventral rod photoreceptors are kept in proper number and it regulates the cell cycle. Cell cycle you know that we have seen in the G1, S, G2 and M phase that it is a regulator during the retinal development, because you are talking about the formation of the retina at the embryonic stage.

So, it has to control the rate of cell cycle of the progenitors and it has to maintain a healthy and adequate number of the photoreceptor cells. Too much also problem, too less also is a problem; it should be just adequate. Another gene is six3OS that is transcribed from the homeo domain transcription factor six3 gene.

So, it is involved in the regulating the retinal cell specification. Six3OS regulates the six3 activity in the developing retina by binding directly to ezh2; ezh2 is a catalytic component of the PRC2 complex. And eya, eya is another gene that is acting as a molecular scaffold for the recruitment of histone modification enzymes to the six3 target site.

So, it can affect six3OS can affect the ezh2 and eya family of genes that can cause epigenetic changes. So, eya is a short form for I absent gene and its mammalian homologs are also well established and it encode a protein tyrosine phosphatase that function as transcriptional co-regulators controlling, I field specification. So, any mutations occurring to eya causes absence of eye that is why eye absence.

So, the I field specification this is somewhat similar to the Hox gene function is affected. Then comes RNCR2 and which is also known as myocardial infarction associated transcript MIAT and it negatively regulates the differentiation of Amacrine Interneurons and Muller Glia, but it does not affect the development of other neuronal subtypes such as bipolar interneuron.

Retina has got a bunch of neurons like rods and cones are the photoreceptors and you know Muller Glia is one of the Glial cell types and you have got horizontal Amacrine and bipolar cells and lastly you have the ganglion cells; so, this forms the retinal structure. So, the formation of specific cell types such as Amacrine Interneuron, it can negatively regulate the differentiation of the Amacrine Interneuron's and the Muller Glia.

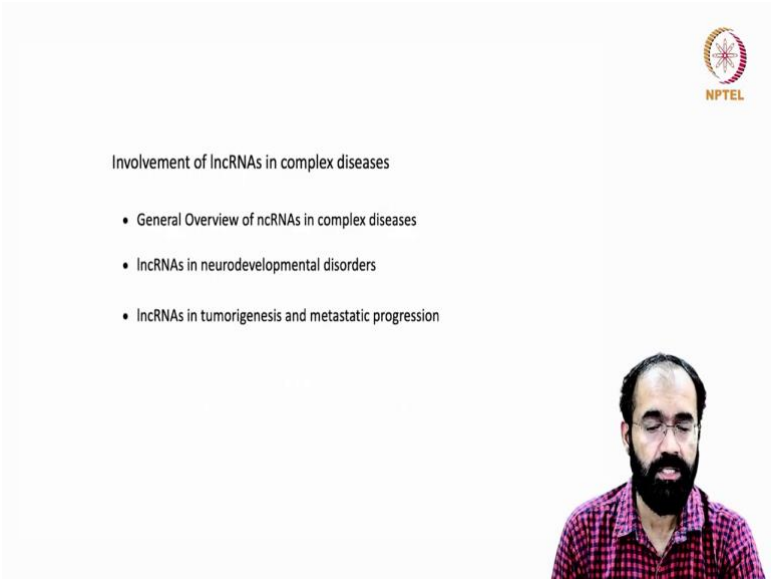
So, but it will not cause any problem to other neuronal subtypes such as bipolar interneuron. So, it is a facilitator of the bipolar cells, but it is a non-facilitator or a blocker of the Amacrine and the Muller Glia cell types. So, it can be a probable target of the transcription factor of Oct4 in MESC cells Mouse Embryonic Stem Cells maintaining Oct4 expression in a feedback loop to help to maintain the pluripotency.

So, if pluripotency has to be maintained a lot of genes has to come into picture, Oct4 is one of the pluripotency factors. So, it is a target of the transcription factor Oct4; so, this is important for the maintenance of the pluripotency. So, the variants it can genes can also have lots of variants though the variants confer the susceptibility to myocardial infarction and that is why the gene has got another name MIAT myocardial infarction associated transcript and possibly through altering RNAs protein binding properties.

So, what you see here a non-coding RNA important for eye field development also have a link with myocardial infarction which is the indicator of heart failure or heart performance failure. So, this also can have a different role, we have seen many such examples one given RNA presence of an RNA can influence one way in one tissue and another way in another tissue.

We also saw in bacteria how sigma factor and the RNA polymerase interaction is influenced by 6S rrRNA if it is present. So, how it is getting affected? Its target selection will change like that many times the presence of non-coding RNA in a unwanted tissue can also have a altered fate determination for that particular tissue type.

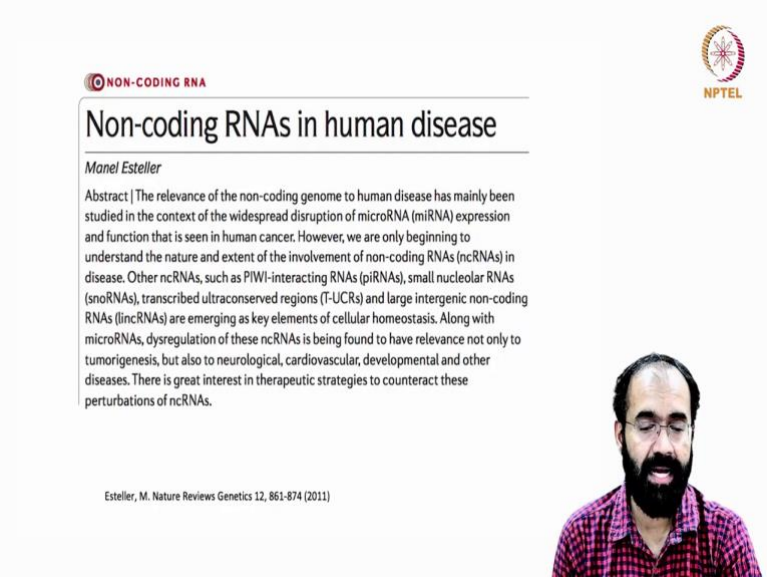
(Refer Slide Time: 19:24)



So, involvement of lncRNAs in complex diseases are a growing field of research and lot of research is going on in that area. So, general overview of non-coding RNA in complex disease is an active area of research and lnc RNAs in neuro-developmental disorders is another active area of research that is going on.

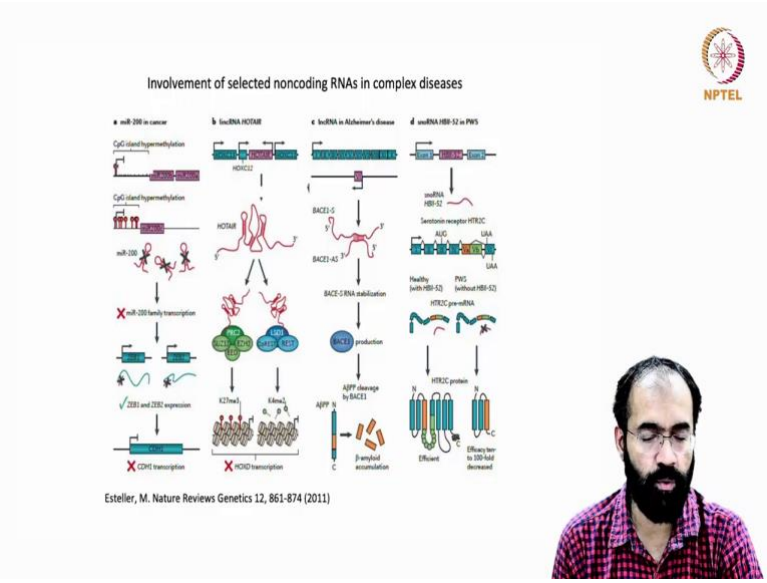
And lnc RNAs in tumorigenesis and metastasis progression is also an active area of research. And we know a lot about the role of non-coding RNAs in complex diseases the neuro-developmental disorders and also in cancer and metastasis, but the research area continue to flourish.

(Refer Slide Time: 20:13)



And they say article that is titled non-coding RNAs in human disease which is a nature reviews genetics article published few years ago. So, you can read it that is a good collection of information about the subject on non-coding RNAs in human diseases.

(Refer Slide Time: 20:36)



So, now let us see in this cartoon it is a it is a dense panel, but you can see the involvement of selected non-coding RNAs in complex disorders. Like, you have miR200 in cancer, lnc RNA HOTAIR and lncRNAs in Alzheimer's disease and snoRNA in Prader Willi Syndrome, PWS. So, now let us see how is it working; so, in miR200 how is it influencing the cancer.

So, there is a area in the gene regulatory regions that is called CpG islands C stands for cytosine, P stands for phosphodiester backbone and G stands for guanine or guanosine. So, if this area is methylated; so, the cytosine residues can get methylated guanosine is not getting methylated the cytosine residues are methylated in the CpG island.

CPG island basically means you have repeats of CG, CG, CG like that; so, or more abundance of C, P and G, P stands for the backbone. So, C pairing phosphor diester backbone, and then G not C pairing with the G; so, do not get confused with that. So, if you have too much of the methylation in the C's of the CpG island the gene expression is negatively regulated means gene expression is not preferred if the methylation occurs.

So, if you have too much of methylation CpG island hypermethylation then the miR200 expression is compromised like you can see here miR200B and miR200A stays like a cluster and it is promoter. If there is minimal methylation the genes will be non-coding RNAs will be expressed, but if it is hypermethylated then the miR200A and B will not be produced it is not available.

Then what happen miR200 family transcription where they are regulating, they will be unaffected they can go smoothly, because there is no negative regulation on those m RNAs where the miR200 is targeting. One example include ZEB1 and ZEB2; so, these are important in the epithelial to mesenchymal transition; so, they are allowed to function perfectly.

And what will happen if ZEB1 and ZEB2 expression happens smoothly, then it can lead to CDH1 Cadherin-1 expression it can influence the expression of the Cadherin-1 in a negative manner. And they will be missing that can lead to the metastasis or the localization of this mad cells or cancer cells into different loci.

So, I will revisit once again if the there is hypermethylation of the regulatory region miR RNAs not expressed. And because of miR RNA not expressed there will be smooth

expression of this ZEB1 and ZEB2 protein. And ZEB1 and ZEB2 proteins are influencing in a negative manner, the CDH gene and that can facilitate the migration of cancer cells or causing the metastasis.

Similarly, lnc RNA hot air if it is not like we know hot air is negatively regulating hot hox D while allowing the hox C clusters. So, if hot air is missing what will happen that the recruitment of PRC2 and LSD1 complex to specific gene expression loci of hox D is prevented. So, you will now end up getting hox D expressed in the absence of hot air; so, because of lack of this epigenetic modification.

Because, hot air is supposed to bring in this epigenetic modifiers to the hox D locus absence of which they are allowed to express. And now you have the lnc RNA in Alzheimer's disease; now, the gene that is studied is BACE 1 BACE1 S. So, it can have a sense RNA and an antisense RNA; so, you have BACE 1 S sense strand and BACE 1 AS antisense.

So, when they are both are expressed that will cause the inhibition of this RNA BACE 1 RNA stabilization will take place if this pairing did not happen the absence of the antisense RNA can cause the stabilization of BACE 1 RNA. And the BACE 1 protein will be produced and this can lead to amyloid plaque formation ABPP cleavage by BACE 1 takes place.

And you can lead to it can lead to the amyloid accumulation in the cell and eventually the organism will end up having Alzheimer’s disease. Sameway if you are seeing the snow RNA, HB2 H2 in Prader Willi syndrome; so, what happens the snow RNA Hb H2 if it is expressed smoothly what will happen? It can affect the serotonin receptor Ht R2C.

So, this can influence in turn the behaviour, because Prader Willi syndrome is a behavior characteristics. So, it can negatively influence the serotonin receptor and it can serotonin is important in controlling the mood and you know depression etcetera. So, the availability of the serotonin receptor STR2C can influence the maturation whether the receptor is complete or incomplete.

This can affect a full length STR2C protein or incomplete protein, why? Because this non coding RNA can affect whether this protein is allowed to translate fully or RNA is allowed to translate fully or partially. So, what we should understand from here is that

different types of non coding RNA employ different strategy to block the gene expression events and drive the cell in a unique manner. So, we will study more in detail about the non coding RNAs in the next class.

Thank you.