Introduction to Developmental Biology

Prof. Subramaniam K

Department of Biotechnology

Indian Institute of Technology, Madras

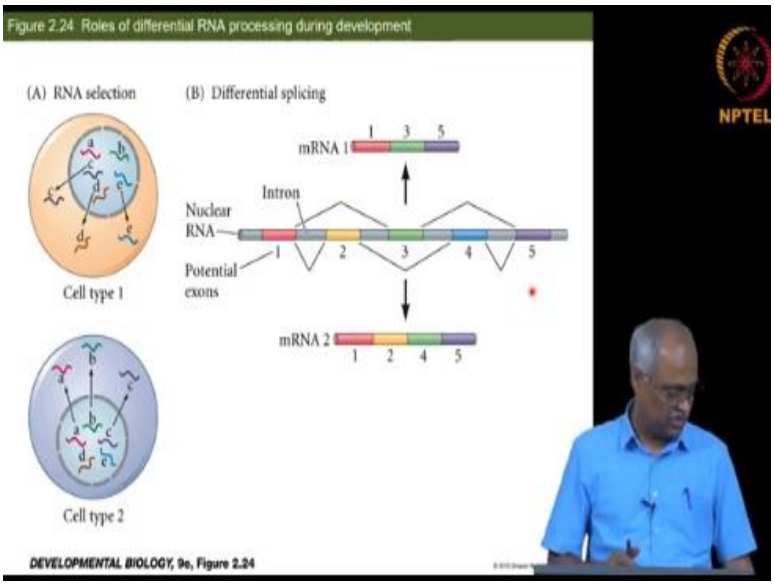
Lecture No-07

Differential Gene Expression (Part 4 of 4)

(Refer Slide Time: 00:14)

So today, we are moving forward from the chromatin to the next stage in the gene expression. That is nuclear RNA processing. So, this is eukaryote specific. Since we are talking about multicellular organism splicing, nuclear RNA processing or other processing are essential.

(Refer Slide Time: 00:51)

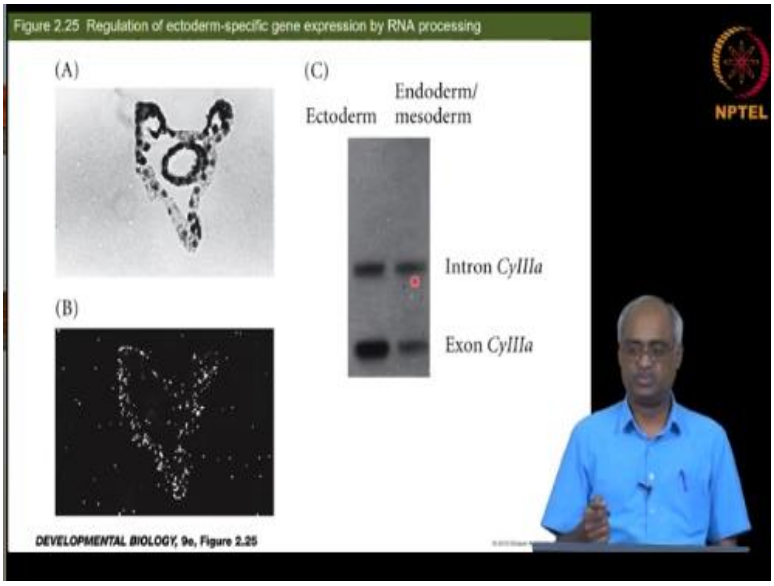


So, in the above slide, if you look at A, there are two different cell types, and in both the cell types, you have five different messenger RNAs (mRNA) being transcribed from five different genes. In cell type 1, mRNA C, D, E enters into the cytoplasm, but in cell type 2, A, B, C come out to the cytoplasm. So this is a nuclear selection that decides which mRNA should be exported and not exported. So even at that level, you could have differences. There are specific developmental contexts where regulations at these steps become important, in the overall context of the embryonic development. We cannot always do it at the level of histone or DNA methylation and controlling transcription. Because there may not be enough time for it, and these transcripts may be used later in some other stage. Therefore, they are made in advance, so that is why these regulations at the post transcription step are important depending on the time and spatial context, and one of them is nuclear selection. The other regulation at the nuclear level on RNA is the splicing.

So, if you look at B, the central bar diagram tells you the actual gene structure as you see it in the nascent RNA just transcribed, not spliced. So, you have 1,2,3,4,5 different exons here, and these V-shaped vertical lines indicate which is going to be joined. For example, the lines on top of the central bar diagram indicate that the exon 1 and exon 3 will be joined, and exon 2 will not be joined, so the exon 2 will be treated as an intron in that case.

So, this splice donor, where I am pointing, is connected to this splice acceptor here. So, skipping two other potential ones in the middle, so that is one way of splicing so you are bringing together exons 1, 3, 5 and in another one you find 1, 2, 4, 5, so this is alternative splicing. So alternative splicing is not a minor component in gene expression regulation, so as you will see in subsequent slides now. So that is where we are going to spend the bulk of today's time.

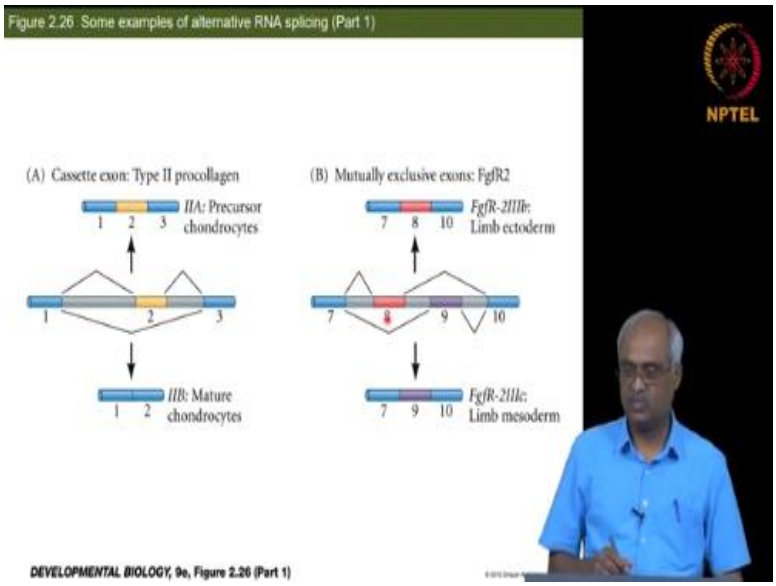
(Refer Slide Time: 03:46)



So, this is the real example of the nuclear selection, so this is Sea urchin embryo at an early stage. So, in (A)You are looking at a phase-contrast image showing all the cells. So, in (B), this is an in-situ hybridization revealing the localization of a certain messenger RNA, which here is CyIIIa. So, it is expressed only in a few ectodermal cells not all over the embryo and, that is further illustrated in this northern blot (C). Here the RNA isolated from the ectoderm in lane 1 and RNA isolated from endoderm and mesoderm in lane 2 are separated electrophoretically and transferred to a membrane and probed with the radiolabeled probe for CyIIIa, intron specific probe that reveals the larger RNA. If you look at the image, the difference is not much, they have a more or less similar level of presence, indicating that in all the three layers it is being transcribed. It is there in ectoderm and as well as in the endoderm and mesoderm. But if you look at the same transcript using an exon specific probe, which is going to hybridize to in a large fraction of the RNA, there will be mature mRNA present in the cytoplasm. Because the processing happens quickly, and they come to the cytoplasm. Therefore, it is largely hybridizing to the mature cytoplasmic RNA. And its level is very high in ectoderm compared to the other

two indicating that there is a selection of the RNA in the cytoplasm like what is getting exported. So, in the ectoderm, it gets preferentially exported but not in the other two germ layers.

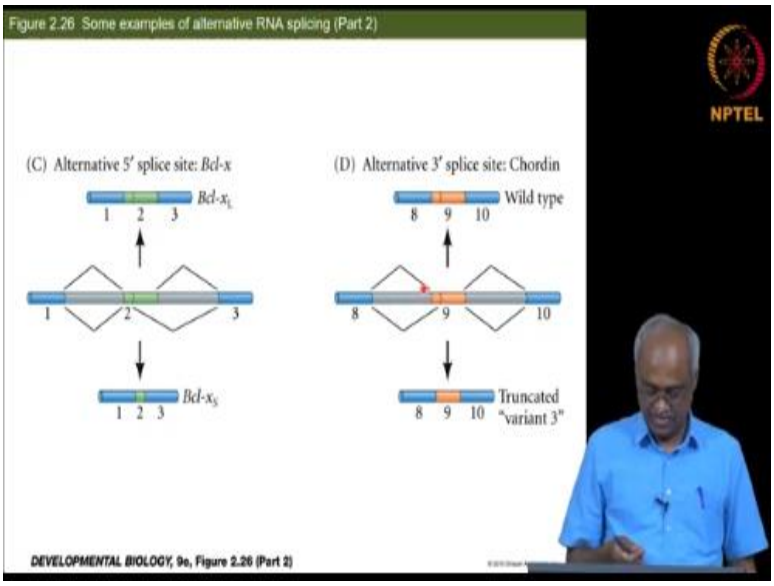
(Refer Slide Time: 06:02)



So, we will continue with alternative splicing. There are four different ways of alternative splicing. The first type is the cassette exon. For example, in (A), the type II procollagen mRNA is present has one cassette with exon 1,2,3 in precursor chondrocytes and as another cassette with 1 and 2 in mature chondrocytes.

The second type is mutually exclusive exons where from a pre-mRNA after splicing one exon will be included in one tissue type and excluded in other tissue types. So here in (B) if red is present purple is excluded and if purple is present red is exclusive, they both are mutually exclusive.

(Refer Slide Time: 07:03)



And the third type in alternative splicing is 5' splice site variation. In (C), from pre-mRNA, exon 2 is fully included in Bcl-x_L but in the other one presence of a splice donor at the beginning of the exon 2 results in excluding the majority of exon 2 and forms Bcl-x_s. So, two v-shaped lines show the beginning of an intron in both the types. So, in this case this has serious consequences. Bcl-x_L inhibits programmed cell death and Bcl-x_s induce cell death. It has opposite consequences and, in most cancer, cell lines, the x_L version is common; as a result, apoptosis is inhibited and the cell continues to proliferate. The fourth type is similar to the previous type but with the 3' splicing. So, in (D) because of the presence of 3' splice donor two different variants are produced. So, these are the four mechanisms by which alternative splicing can happen.

These types of splicing happen to almost all genes, more than 90% of the human protein-coding genes are subject to alternative splicing. To give you an idea the number of genes in *C. elegans* and humans are not very different, but if you look at the organism's complexity either anatomical or functional, *Homo sapiens* can learn all of this and come and teach a class but *C. elegans* cannot do that. *C. elegans* has a very simple anatomy with a couple of tubes but then having all the basic biological processes; it has a nervous system; it can sense pain and so on. So, the bulk of the complexity between *C. elegans* and *Homo sapiens* comes from alternative splicing. So, therefore, this is a major thing so do not take it as an aberration in the gene expression. We will see some more examples.

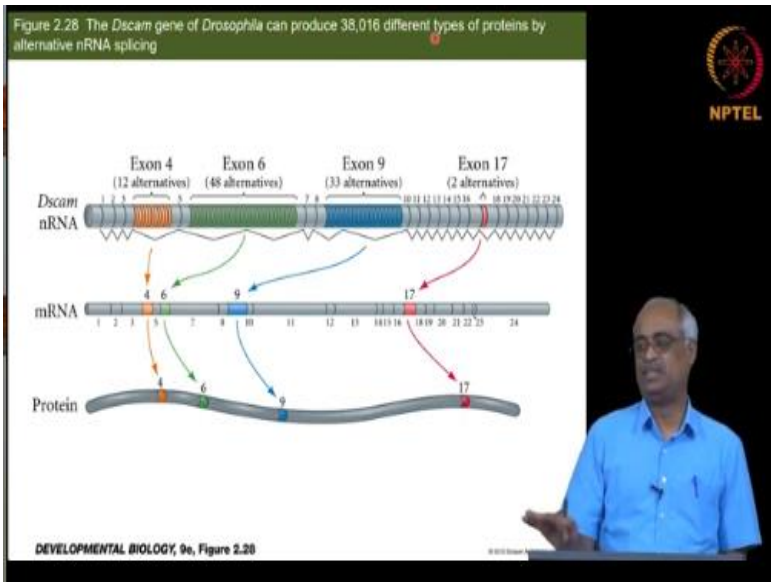
(Refer Slide Time: 10:23)

So, the next example is α-tropomyosin in the rat. In different muscle different exons are included. So, these green colored ones are exons present in all the different variants and another important thing is sometimes alternative splicing can affect the 3'UTR sequence.

Striated muscle mRNAs have a different 3'UTR from the other muscle mRNAs. So, the importance of 3'UTR will become clearer when we go to translation regulation in the future slides. So here some exons included are smooth muscle-specific and some exons are specific to straited muscle and only they are included in striated muscle's mature mRNA.

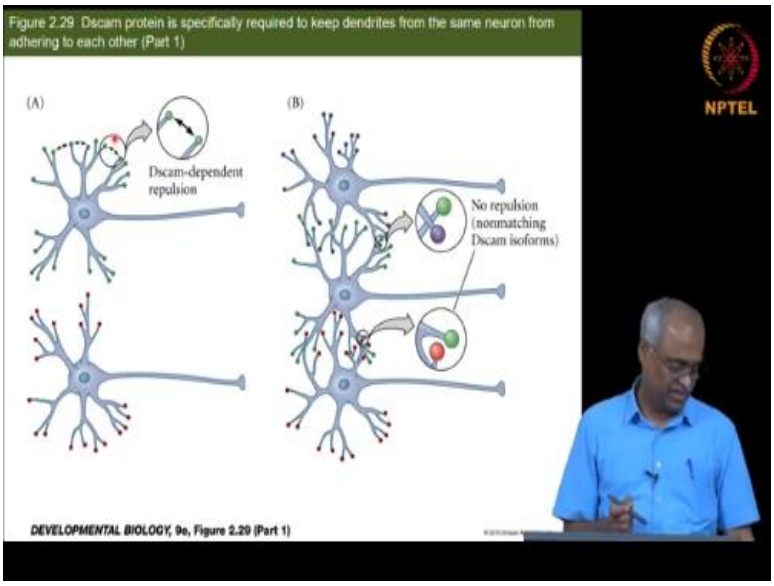
So, by these various permutation combinations of excluding and including exons variety of proteins can be generated. So, therefore, the definition of one gene-one enzyme that became one gene-one polypeptide now actually becomes one gene and one family of proteins. These are protein family because they do have some similarity, some exon sequence is there in all of them.

(Refer Slide Time: 11:59)



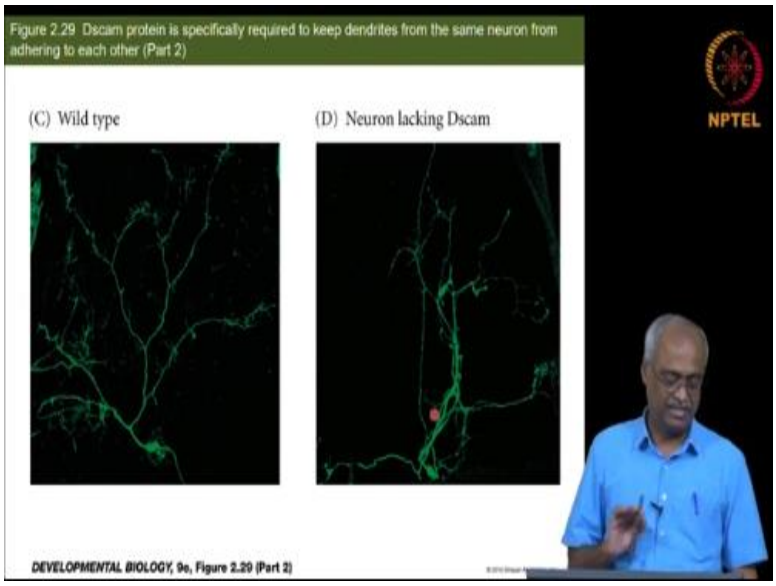
The most complex alternative splicing that we know currently is Dscam, Drosophila neuronal mRNA that is expressed in neurons. We will see the functional importance of this in the next slide, but this illustrates, from one single gene you get 38,016 different mRNAs that are nearly double that of the entire protein-coding genes in Drosophila. Drosophila has only about 14,000 genes so it is nearly double of *C. elegans* genome itself. So here 24 different exons are present in this protein. From the nascent mRNA in the striped segments, anyone can be selected as Exon 4; 12 different alternatives are possible to serve as exon 4, 12 different adjacent sequences. Similarly, for exon 6, 48 alternatives are possible, and then exon 9, 33, exon 17, 2. Now if you work out all the combinations it works out to be 38,016 and experimental data supports that the vast majority of them are all produced.

(Refer Slide Time: 13:23)



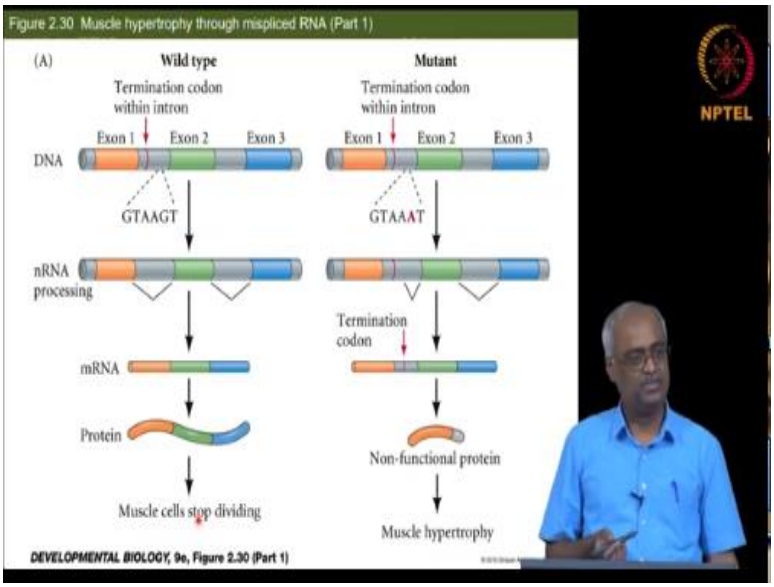
So now let us look at why that is important and how that helps. So, this is more like the VDJ recombination in human b-cells, doing a similar purpose here but for self-recognition. So here in the neuron, you have branched dendrites and these dendrites need to know that they belong to the same cell body. So essentially, they are sisters and they need not connect. When they get a signal, they need to connect to the adjacent neurons or another neuron in that chain then only it will be useful for nerve impulse conduction. So, this self-repulsion is provided by that unique Dscam alternate spliced version produced in that neuron and another neuron has different alternative splicing. Therefore, it has its Pincode in terms of Dscam exon composition and that helps in connections. In (B) green and purple form one connection and green and red form another. Suppose let us say this Dscam is mutated and it is not produced so what would you expect? These will not know to repel and they will all get adhered and they will make a mass of connected neuronal structure.

(Refer Slide Time: 15:03)



So that is what happens. So, the (C) wild-type is highly branched and (D) the neuron lacking the Dscam. So, this is alternative splicing.

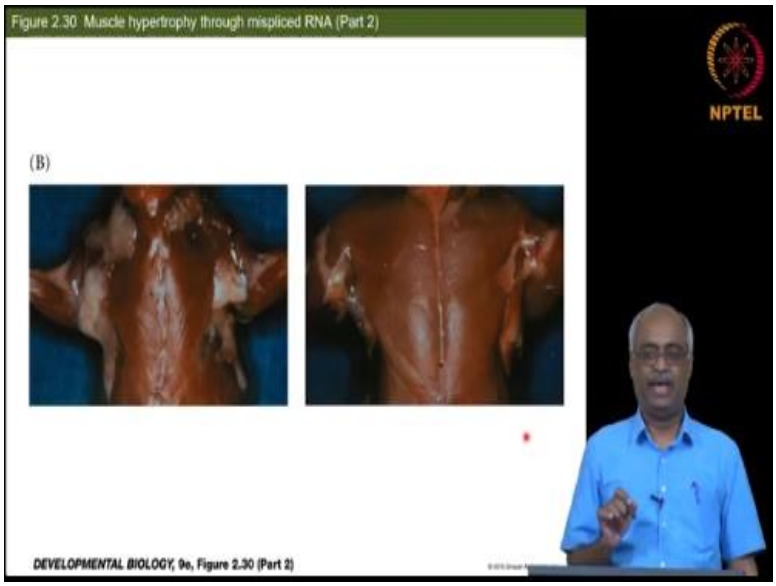
(Refer Slide Time: 16:00)



Another interesting example is the muscle-specific splicing. This work happened exactly around the time I joined the department as a postdoc, and this paper had come out from another lab in the same building. It was in the press, and it was talked about, so I was excited to read, and that is there in the textbook, so we will read about this, so here you have a muscle-specific splicing issue. Now we are going to look at disease conditions, so within Drosophila, we saw we can mutate and find the function. So, in human patients, is there a consequence when these splicing do not happen as they are supposed to be happening? So, we will see an example.

In (A), between exon 1 and 2 a stop codon is present. During splicing, this intron is excluded, so a functional protein is made. But in the mutant, you have a mutation in the intron where a G becomes an A and due to that it promotes splicing. So therefore, a part of this intron is included, and due to that, translation is terminated and you do not make a functional protein. So, in wild-type, that stop codon would never be encountered because that whole thing is excluded, and as a result, you straight go to exon 2, and the reading frame continues to make the wild-type protein. So, the name of this protein is myostatin. This protein prevents muscle cell proliferation at a specific time point, and they differentiate into muscles right away, which helps in making the required amount of muscle. So, in a family where they found this mutation, four guys were already athletes, and a small kid was able to hold a 3 kg dumbbell with both hands fully extended. This was possible because their muscles were so strong, and this can be done in model organisms too.

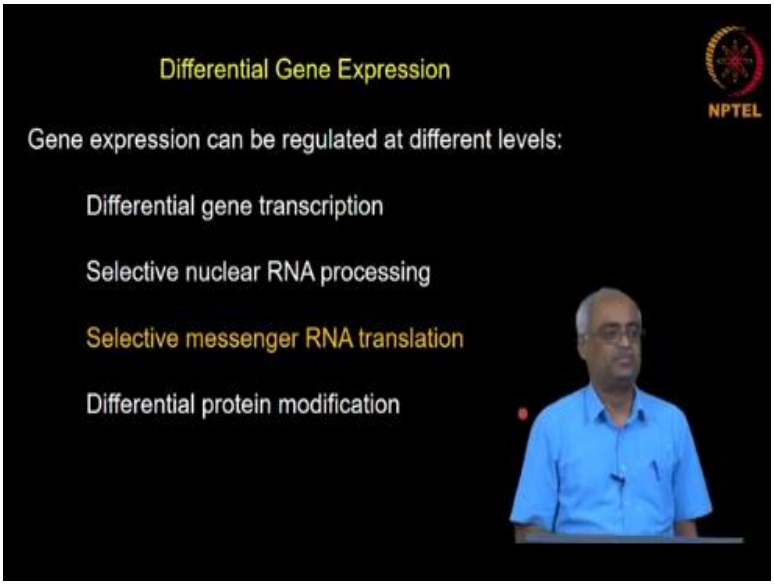
(Refer Slide Time: 19:06)



So, in this slide, the one on the right side is the mighty mouse, the left one is a normal mouse. This attracted press, the paper came in nature, and then this faculty was on the press to meet with TV, radio, and newspapers. So, you see the muscle difference between the two because the cell proliferation was not stopped at the right time, so they made a lot more cells, and when they differentiated, they made a lot of muscles.

So, will a developmental biologist right away recommend to a physician that you mutate and make more muscles? Will not because developmental biologists will think about the other consequences. For example, you will worry about methylation patterns when you think of only the coding sequence. So similarly, you will expect there will be other complexities like, for example, a phylogenetic and ontogenetic reason for why that cell division stopped at that time point. So, you will not immediately mess up, but this gives you an idea of what happens when you have a splicing defect.

(Refer Slide Time: 20:29)

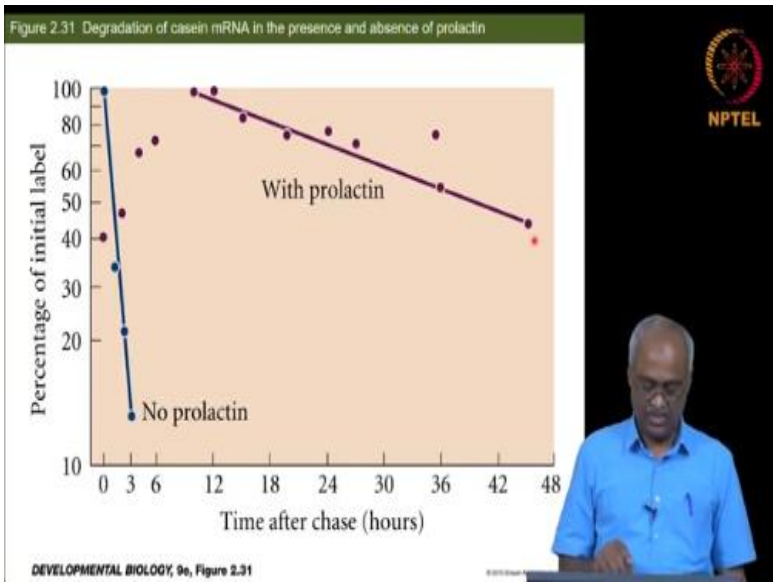


I think that conveyed the message about the importance of splicing. So next, we will move on to translation regulation. So, translation regulation is very vital in certain contexts; there are two examples outstanding examples to illustrate this point; one is early embryogenesis, and another one is the adult nervous system. You take a neuron; for instance, it has a cell body and has a long axon, and at the end of the axon, you have synapses. At the synapse, neurotransmitters are produced and secreted. Then the postsynaptic neuron must have the receptor to accept the signal. If these things have to be made and degraded very quickly to sync with the nerve impulse conduction, you cannot go and unwind the DNA by modifying the chromatin and then transcribe, splice, and then get permission from the nucleus to come out.

Neurons will not have enough time for it and after that, the protein has to be transported to the end of the axon to get to the site of synapses. So that problem is usually solved by making the RNA and keeping it stored where you want and when the correct signal comes, you translate. That is one context; the other context is the cleavage in the embryo. It is rapid, and during embryogenesis, new transcription is absent. So, to carry out the embryonic development, particularly coordinating the cell cycle and producing a lot of cell membrane, required factors should be already made and stored in the oocyte. Also, during cell division, a lot of nuclei are made, so we need that much of a chromosome. Then the cell division cycle needs to be regulated. Once you cleavage is completed in some organisms, at that time, cell fate specification happens; for example, in *C. elegans* at the first division itself, the fate of the two daughter cells is different.

So, all of that cannot be controlled at the transcriptional level. So there again, translation regulation of mRNA produced by the mother and deposited in the oocyte plays a critical role. So much so that in some organisms, you can get rid of the nuclear components, but still, the cell cycle can happen without the nuclear material; the cytoplasm has the required things to simulate that, and these are contexts where translation becomes important. So now, let us see how the translation regulation happens.

(Refer Slide Time: 23:58)

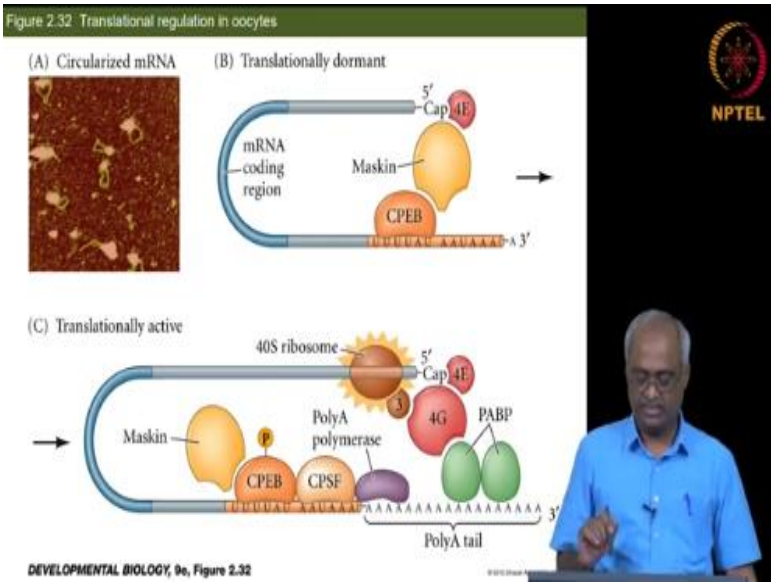


So, one mechanism by which translation regulation can happen is increasing or decreasing the stability of the messenger RNA, meaning once transcribed, how long will that messenger RNA survive before getting degraded. So, you can have a half-life for that, the duration taken for half of the messenger RNA to be degraded. So, in a lactating mother, the prolactin mRNA generally produced in the mammary gland gets quickly degraded. I should explain the method first. So, this is mouse mammary gland cells cultured, and incubated with radioactive nucleotides like in this case, probably UTP. So, all the RNA synthesized gets radioactively-labeled. Then cells of that medium containing radioactive nucleotide are washed and put in a fresh medium without radioactive nucleotide but you have unlabeled one, so that is called the chase period.

Now you watch how long does the labelled RNA stays. So, without this lactation stimulating hormone prolactin the mRNA gets rapidly degraded, half-life is probably about two hours or one-and-a-half hours but with the hormone it stays for more than two days. So, by increasing the stability the mRNA is available for multiple rounds of translation, so you increase the protein product.

So, the ultimate goal of differential gene expression is to vary the kind and amount of proteins from one cell type to another cell type. So that is the ultimate end point of gene expression when you talk about protein coding genes. So, this is an example for differential stability, so here the stability is under two different conditions.

(Refer Slide Time: 26:19)



The next one that we see is what happens typically in oocytes. This was first described in good detail in Xenopus oocytes; therefore, we are going to look at that example, and this is the kind of regulation that is critical in almost all the oocyte development that people study. So, in fact, in our lab, we exclusively study translational regulation of the germline. So, let us look at this; before that, I should introduce you to this idea of looking at the mRNA.

So usually mRNA is written as a long line with 5' cap to 3' poly-A tail, but in cells, they do not exist like that; instead, they are found in the ring-like form. So, the 5' end and 3' end are brought together by proteins that bind to the 5' cap, for example, translation initiation factors that bind to the 5' cap or to 5'UTR etc. They interact with proteins that bind to the 3' UTR or poly-A tail, and those protein interactions make the mRNA into a circular structure.

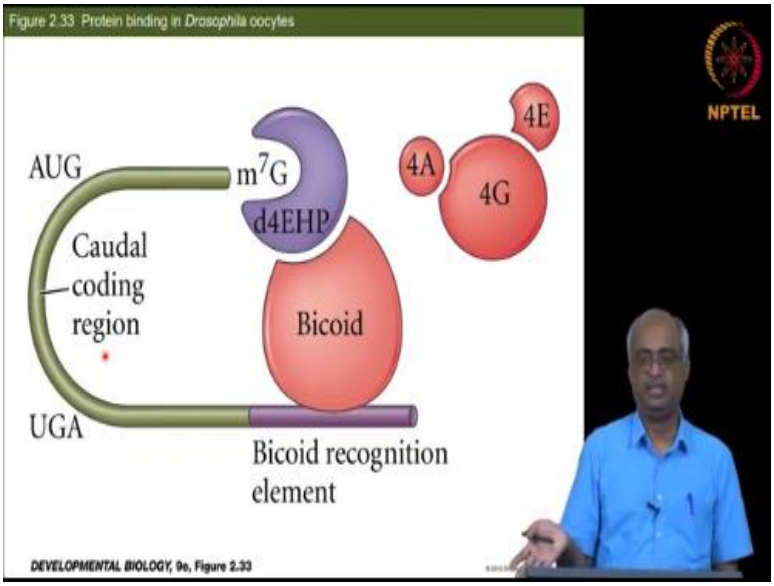
So that is what you see in (A) an electron micrograph. So here is one example (B), so usually the initiation factors binding to the 5' cap interact with a protein that binds to the poly-A tail, called poly-A binding protein (PABP), which results in circularization. And that is required for binding of another initiation factor-like 4G, which recruits the small ribosomal subunit and initiates translation. So that is what happens typically but in Xenopus oocyte, the most mRNA that is not translated at a given time are found in circular form but not by the interaction of 4E and 4G with the PABP; instead, a protein called Maskin binds to 4E and that Maskin interacts with a protein called cytoplasmic polyadenylation element-binding protein, CPEB that binds to the

polyadenylation signal sequence. Now CPEB interacts with Maskin, and this kind of circular form totally excludes the other initiation factors, and therefore, it is translationally dormant. It does not get translated until appropriate signaling comes. For example, progesterone hormone, which stimulates oocyte maturation and completes the mitotic division upon fertilization, leads to activation of a kinase that phosphorylates the CPEB. And when the CPEB is phosphorylated, it recruits a protein called cleavage and polyadenylation specific factor, and that recruits a poly (A) polymerase. So, this poly (A) polymerase extends the poly-A tail in the cytoplasm. Remember, when we learned about the eukaryotic gene structure, I said some polyadenylation happens in the nucleus. In some of the developmentally regulated one's additional polyadenylation happens in the cytoplasm, so this is an example of that. So, you have a further extension of the poly-A tail, and this poly-A tail now binds the PABP, and that can directly interact with 4G, and this phosphorylation also throws the Maskin out of interaction with the 4E, and now translation starts. Though this has not been totally worked out like how this is choreographed, this gives you a mechanism of how activation happens.

But then there is a wave of sequential activation. Some mRNA is expressed at a certain stage, and some mRNA is expressed later and so on. So that is still being intensely studied, so this whole mechanism was worked out at the beginning of 2000, you know like the papers were coming out in 2001, 2002 around that time. So now, this mechanism has been found in many other organisms where you can do a lot of genetics.

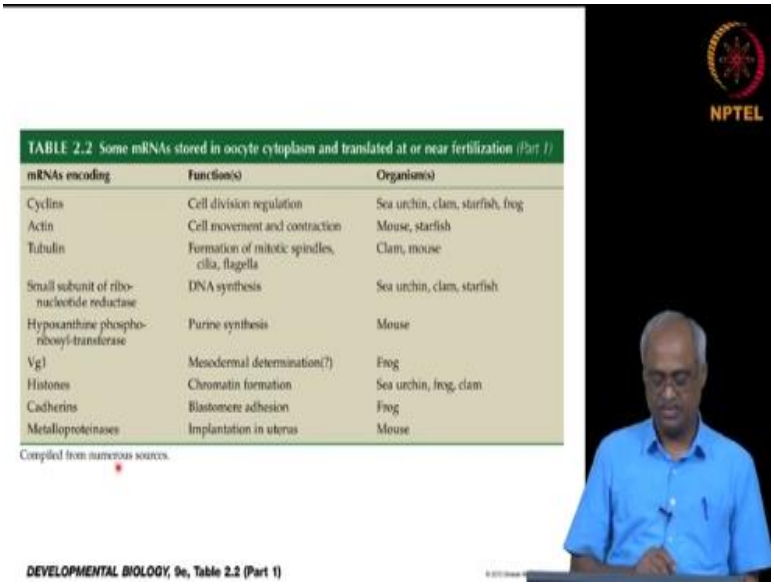
So, there they have done a whole lot more work on many examples of this kind. So, let us look at one such example.

(Refer Slide Time: 31:18)



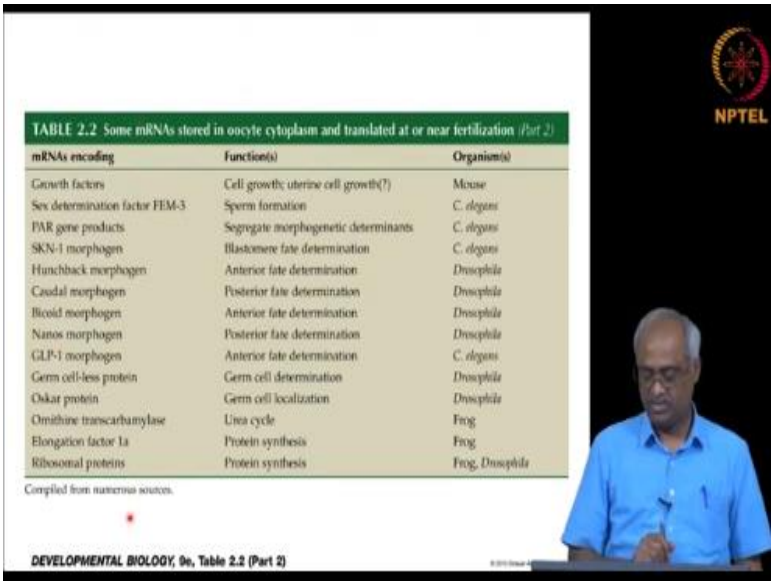
So, this is the Drosophila oocyte, where you have a protein called a Bicoid. So, this Bicoid is an interesting protein; it is a transcription regulator as well as a translation regulator. Here we see it as a translational regulator. So, it binds to a sequence called Bicoid recognition element on an mRNA that encodes a protein called Caudal. So, this Caudal is required to prevent the posterior development in the anterior region. So Caudal needs to be suppressed, and that is done by a Bicoid, which generally promotes anterior development. When Bicoid binds to this mRNA, it recruits another new protein that binds to the methyl cap, and this excludes all the initiation factors. So, this is how Bicoid prevents the translation of Caudal mRNA. So now you realize that the general theme here is proteins that interact with RNA sequence can interfere with translational activation. Some of them increase or decrease the poly-A tail in some contexts, not in the Xenopus oocyte context; where even with shorter poly-A tail, the mRNA is stable, it is just that they cannot bind PABP, and therefore, they are not translated, but they are not degraded. In some instances, if you remove the poly-A tail, the RNA is marked for degradation.

(Refer Slide Time: 32:58)



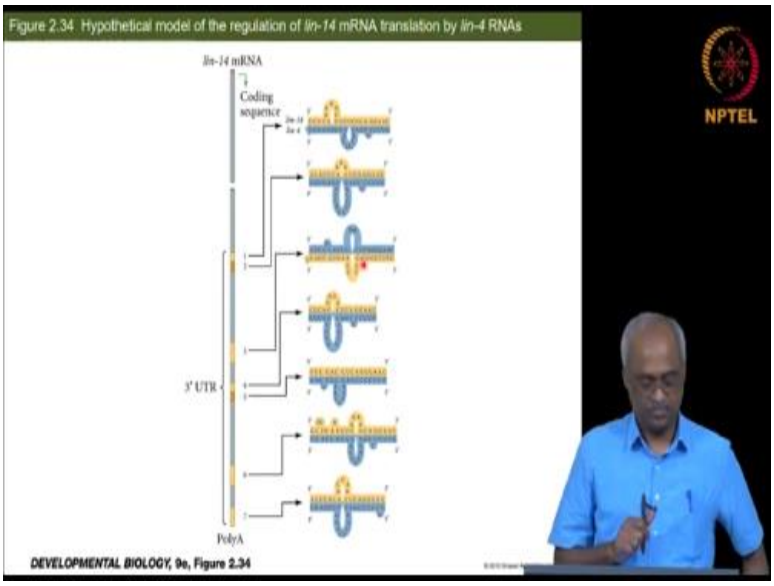
So here, there are some examples of the mRNA molecules that are regulated, where they function, and in what organisms they come from. Cyclin being in the list is not surprising because it is involved in cell cycle regulation.

(Refer Slide Time: 33:17)



So, it goes on, so from multiple organisms, so this is an evolutionarily well-conserved mechanism.

(Refer Slide Time: 33:27)



Next, one is obvious, but then it took a long time for people to find an example of this. If proteins can bind sequence specifically to 3' UTR and 5' UTR, why not another nucleic acid? The sequence complementarity should give you more specificity to the given sequence, and that was found in an exciting screen in *C. elegans* embryogenesis. So, in *C. elegans* embryogenesis, a specific pattern of division can end up reiterating if specific gene expression changes do not happen.

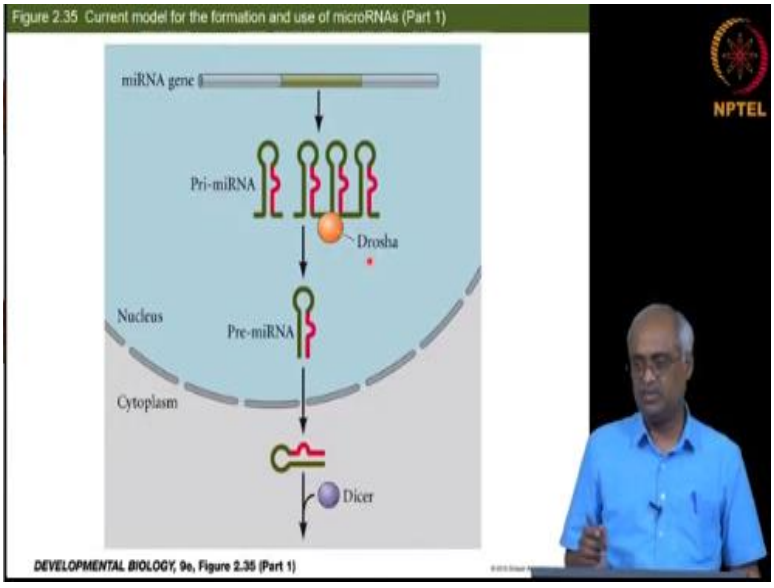
Let us say, at the fourth cell division level, a specific pattern of asymmetry is generated during cell division, and you do not want that to be reiterated because you will end up producing similar kind of cell fates in another generation as well and that is avoided by blocking some gene expression. So, mutations in them are called heterochronic because, in the time order, they are at the wrong time. So, what should have been at the third stage will end up repeating at the fourth stage or the fifth stage or even later, and therefore they are called heterochronic mutations. The lab that identified the mutation that had the heterochronic phenotype finally mapped the mutation to a locus, but they did not find any proper protein-coding sequence there. They were confident about the phenotype and the locus they mapped, and they knew that locus matters even if it is not encoding proteins. When they carefully looked at, they found it encodes a short RNA, a small RNA sequence that is complementary to the 3' UTR of another RNA that is involved in this lineage specification. So *lin* strands for lineage in *C. elegans*, most genes have three alphabets and hyphen and a number.

The number usually is the order in which that particular gene belonging to that class was identified. So *lin* meaning lineage defective, in that it is the fourteenth gene that they identified. So, if you look at *lin-14* mRNA sequence in the 3' UTR, there are colored sections 1 to 7, and those are the sequences to which this *lin-4* binds. *lin-4* is one of the heterochronic mutations which was not encoding a protein and found to be a small RNA.

The loops indicate where there is no complementary, where you see a linear thing; there is base matching. So, this was the first discovery, and is this something nematode specific? NO, in the last twenty years, people have discovered this small RNA is naturally encoded in the genome just like the *C. elegans* genome, and they regulate the translation of the vast majority of mRNAs.

The current estimate is about 30 to 40% of all human mRNA are subject to regulation by these small mRNAs called miRNAs, micro interfering RNA molecules, or micro RNA shortly.

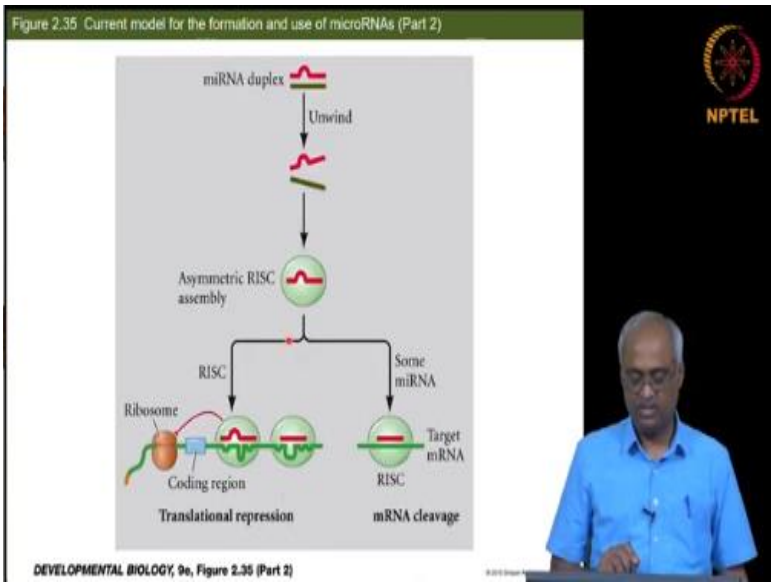
(Refer Slide Time: 37:10)



And this is how they are made; they usually exist in multiple tandem repeat copies. Once transcribed, a nuclease called Drosha cleaves them. So, the information in this carton comes from studies in various organisms. The first example is from *C. elegans* that is where it was first discovered, but then people worked out the mechanism by studying multiple organisms.

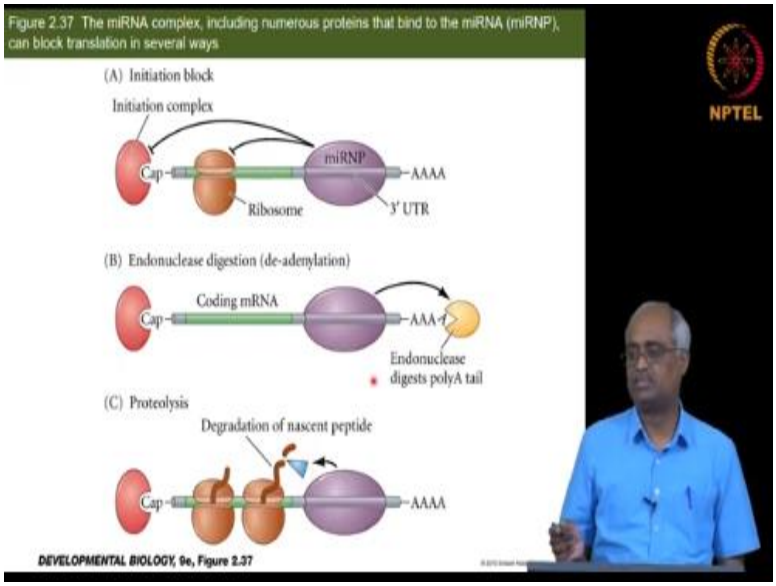
For example, Drosha was discovered in Drosophila, where they made cell extracts and did these reactions in-vitro. So, these multiple repeats are digested into individual units in the nucleus by Drosha, and that is transported into the cytoplasm. These come as hairpin repeats, and that hairpin is removed by a cytoplasmic nuclease called Dicer.

(Refer Slide Time: 38:07)



And now you get the two strands, and they are unwound and loaded on a complex called RISC. This RISC complex takes one of the two strands of the miRNA duplex and uses that to identify the target sequence, goes there, and binds to the target sequence. Binding to the target can have multiple consequences; one of them is the RNA gets cleaved, stability is reduced dramatically, this is one, but you can have various consequences.

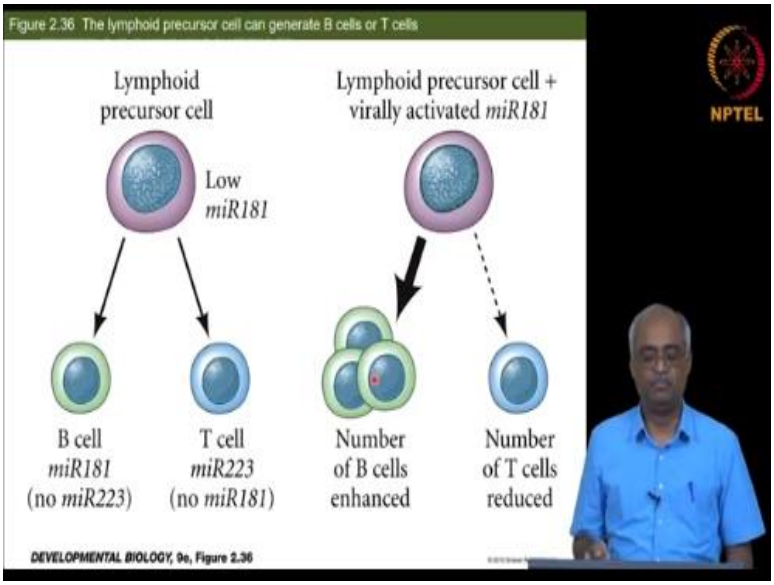
(Refer Slide Time: 38:44)



That is there in this, slide then we will go back to that example. So this miRNP ribonucleoprotein complex carrying the miRNA sequence can perform any of these three; it could inhibit initiation complex, a protein binding to the 5' cap or it can hinder ribosomes ability to elongate, or it can remove the poly-A tail and reduce the stability of the RNA, or it could recruit proteases to digest the nascent peptide that is coming out.

All three mechanisms have been seen with specific examples. So we will see to what extent it matters in the mammalian context.

(Refer Slide Time: 39:35)



So, during the development of different kinds of lymphocytes, the lymphoid precursor cell has a low abundance of miR181. So, you get both B cells and T cells. B cells express miR181 in larger abundance than the precursor. But if you artificially introduce a large quantity of miR181, in this precursor cell, it will end up producing vast majority of B-cell type and not the T-cells. So, you make more B-cells at the expense of T-cells, so these have clear developmental consequences in multiple organisms. So, each cell type is going to have one specific set of gene product, not genetic information, and as a result, its fate is going to be different.