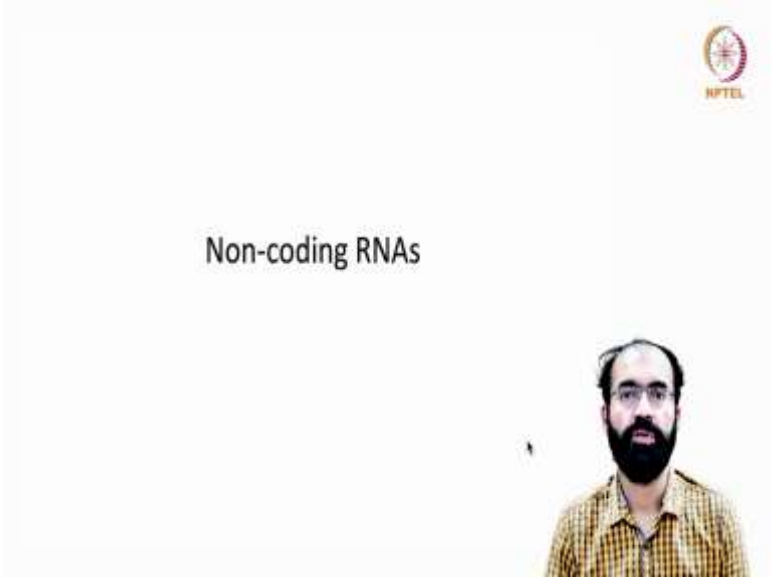
RNA Biology
Prof. Rajesh Ramachandran
Department of Biological Sciences
Indian Institute of Science Education and Research, Mohali

Lecture - 39
Mechanics of RNA Decay and Non Coding RNAs: Introduction to Non-Coding RNAs

(Refer Slide Time: 00:21)

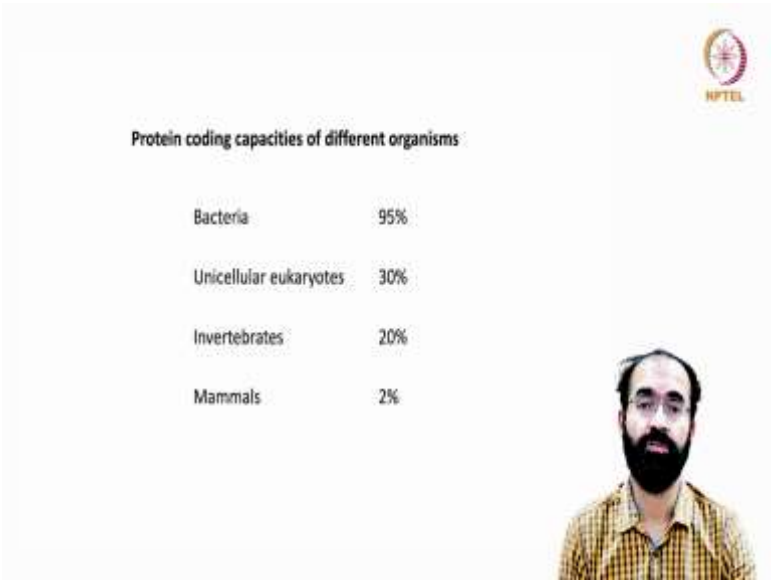


Hello everyone, welcome back to another session of RNA Biology. So, we will move on to a new topic that is the non-coding RNAs and non-coding RNAs are those RNAs that do not code for a protein. And I am sure we have we are familiar with this terminology by now we have used this word non-coding RNA a lot. But the importance of non-coding RNA was not very obvious until recently.

Recently means almost 10, 15 years ago where we started understanding and we are continuing to understand a lot. Because if you see the research articles that are being published in the related field of non-coding RNA, the number is lot does increased the number of publications have increased a lot. And that itself indicates that we are focusing a lot of our research on to the non-coding RNA's.

And non-coding RNA's can have huge implication in the gene expression events, that is what it makes this molecule this non-coding RNA molecule becomes so important in biology.

(Refer Slide Time: 01:46)



So, the protein coding capacities in various organisms vary based on their genome size. We all know bacteria typical bacteria E. coli have got a genome of 4 MB that is 4 megabase or 4 million base both are more or less than the same. So, of this 4 MB around 95 percent of this 4 MB is having protein coding capabilities that means, this does not mean that bacteria do not require non-coding RNA of course, they require ribosomal RNA of course, they require transfer RNA etcetera.

Nonetheless, the protein coding capability are more and they are distributed almost completely the size of the genome itself. In unicellular eukaryotes such as yeast and several what you call several varieties of fungus they are unicellular eukaryotes, means they have a proper nucleus that is why eu means real karyon means nucleus.

So, they are unicellular organism or they are single cellular organisms, but they have a nucleus unlike the bacteria that is why we call it eukaryote. So, in this eukaryote which has a nucleus the 30 percent of the genome only is coding for protein coding capability. So, that is a huge difference 95 percent is dropped down to 30 percent and with a purpose.

And if you look into invertebrates basically means insects like mosquito, fly means they do not have a vertebral column. So, like a cockroach these are all invertebrates butterfly, different types of worms, the model organisms C elegans they are all invertebrates because they do not have a vertebral column.

So, we are all vertebrates like mammals, fishes, birds, you know reptiles they are all vertebrates. So, invertebrates its 20 percent of the total genome is having coding sequence or code for proteins whereas, in mammals like mammals is the one of the most evolved vertebrate group. They have 2 percent of the genome that is coding for proteins.

So, 98 percent is non coding, that itself says that there is a huge potential for having those genes that are not meant for coding amino acid or coding proteins means they are not expressing mRNAs. So, a lot of non coding genes or non-coding RNAs, we are still continuing to explore we do not fully understand.

The reason being a given tissue like you can think about we have plenty of tissue in our body like liver, kidney, heart brain skin many more, you have almost around 100 different types of tissues we have in our body. But which non coding RNA expressed where, at what time is very hard.

So, if you want to fully understand the entire expression pattern, entire genome expression in say human, we need to look into each and every tissue and any given time. We cannot just look say today right now someone is taking a given tissue in my body say liver right now.

So, it may or may not have a given non coding RNA or may or may not have a given gene, unless it is a housekeeping gene. But say if I am facing a bacterial pathogen or I am facing an infection or maybe I am fighting a disease in that situation the same given tissue will have a different expression pattern including that of the non-coding RNA.

So, what we should understand although you have plenty of genes and plenty of non-coding RNA genes they need not necessarily be expressed at any given time. So, understanding them becomes a challenge we do not know what time point a given gene is expressed. We will see some examples.

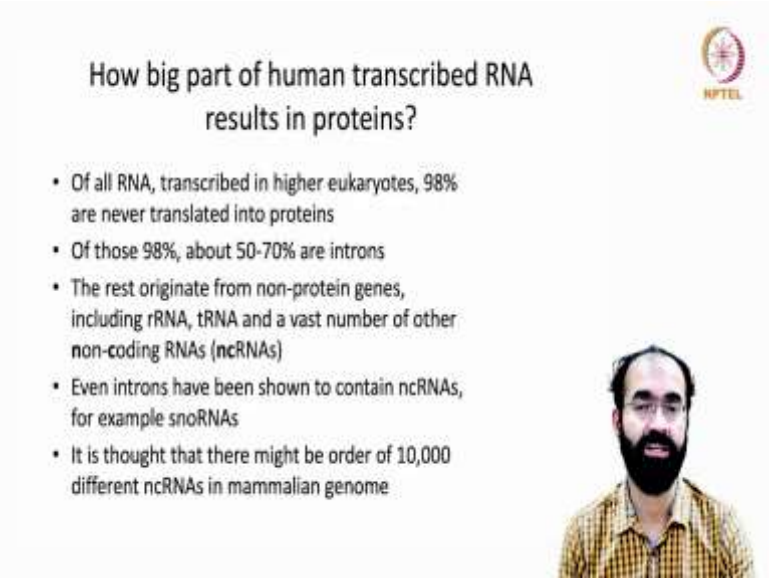
Some genes are simply expressed during a few days during the development when you are developing inside the womb of your mother first few days only, they are expressed after that they are never expressed in your lifetime. So, how are you going to understand?

Of course, there are ways and means of using model organisms, what if a given non-coding RNA is exclusively specific to human being you will not get any human being to

study, except that you may get some aborted fetus, that baby died inside the womb some people can access those tissue and you can study.

But you will not have all time points every day like first day, second day, third day, fourth day any given hour you would not have. So, the understanding of the non-coding RNA will remain enigmatic for a long time to come especially connected to homo sapiens that is human beings it is very very difficult.

(Refer Slide Time: 07:00)



So, how big part of human genome is transcribed RNA that results in proteins? So, how big a part of the human is transcribed RNA that can result? So, let us see of all RNA transcribed in higher eukaryotes, 98 percent are never translated into proteins. Higher eukaryotes means mammals, humans like that they are never translated into proteins. Of these 98 percent about 50 to 70 percent are introns.

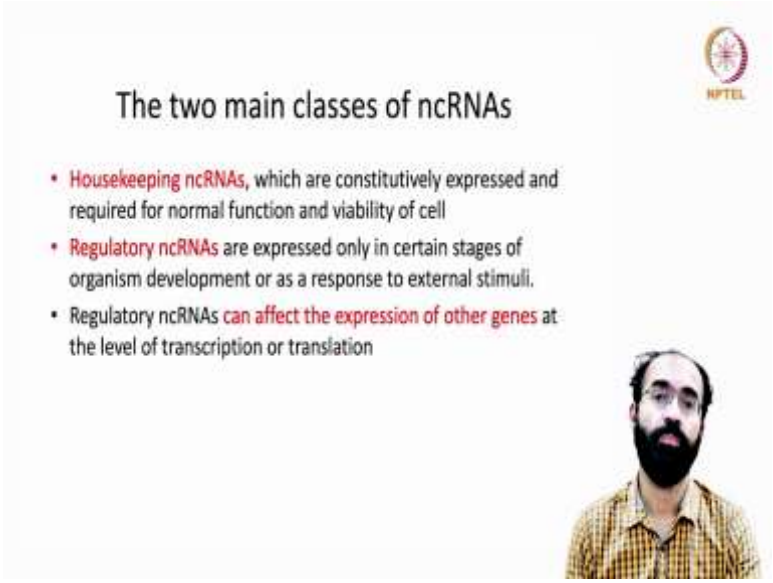
Although they are coding region, although they are coding part of this 98 percent a majority of them are that is around 50 to 70 percent are introns. That means, they are between the exons of a given gene. And introns we know they are huge; they can be really gigantic compared to the size of the exons.

The rest originate from non protein genes, we know some of them ribosomal RNA, transfer RNA etcetera that including ribosomal RNA, transfer RNA and a vast number of other non-coding RNA we call it as ncRNAs. So, they can constitute a significant

proportion. Even introns have been shown to contain non-coding RNAs, some example is snoRNA small nucleolar RNA and several micro RNAs are also encoded inside the introns and they are processed during the transcription of that gene.

It is thought that there might be an order of 10,000 different non-coding RNAs in a typical mammalian genome. 10,000 is a huge number considering we know only 23,000 of gene coding RNA sequences around 23,000 protein coding genes only we know. So, considering that 10,000 is a huge number, that does not mean that there is only 10,000 it can be even more.

(Refer Slide Time: 09:10)



There are two broad classes of non-coding RNAs exist and one class is housekeeping noncoding RNA which are constitutively expressed and required for normal function and viability of a cell. Normal function and viability; that means, that is called housekeeping. A given cell have to do lots of routine functions.

Just like you may be a bank officer, you may be an IAS officer, you may be a student, you may be a politician or you may be a film actor all these professions are different, but there are some things which you will be doing as irrespective of your position. What are they? Sleeping, having food, having tea or coffee, having some water or listening to some songs, watching movies.

So, these are all some of the jobs which you will be doing irrespective of your position or your profession. So, housekeeping genes or housekeeping functions of a cell basically means those jobs a any given cell or any live cell has to perform. Say making use of glucose, it has to break down glucose by glycolysis, then crepe cycle and make ATP. No cell can escape from this duty, it cannot say that ok, I am trying to import ATP from neighborhood. No, that is not available, every cell has to produce its own ATP.

Like that many every cell if a cell has to divide, that given cell has to divide 1 into 2, 2 into 4, 4 into 8 in that situation all genes required for mitosis, then that only cell only has to produce. Like that many such housekeeping genes are expressed which is of the order of around 7000, 7000 to 8000 genes are considered to be housekeeping.

And then comes the second category of genes that is tissue specific genes. So, non-coding RNA also can be considered housekeeping non-coding RNA, they are those non-coding RNA, no matter which cell type you are or no matter which cell type you are studying these non-coding RNAs are expressed.

Then comes the regulatory non-coding RNA name itself says regulatory means they have some regulatory role that means, they are expressed only in certain stages of an organisms development or as a response to external stimuli; that means, they are expressed based on demand. Say, routinely you will have a dress like if you are a male you will be wearing shirt and pant.

But if you are going for a marriage, say your marriage or someone very close to you, you will not go simply in a shirt or and a pant, you will have maybe some occasion associated dress, say some sherwani or something related to that. Why so? Because based on the demand. What is the demand here?

Here there is a marriage you are supposed to be this does not mean that shirt is not a civilized dress, when you go for job you are wearing shirt and going its well accepted. That is even if you go for a high level of a meeting then you will put a jacket and maybe a tie also.

So, like that a given cell also based on demand, based on the situation, based on the circumstance it will turn on some given set of genes. And this in turn is regulated by some of this non-coding RNAs, they are called non-code regulatory non-coding RNA.

You should not think that for a gene expression it always depend on a non-coding RNA, no, that is not the correct assumption.

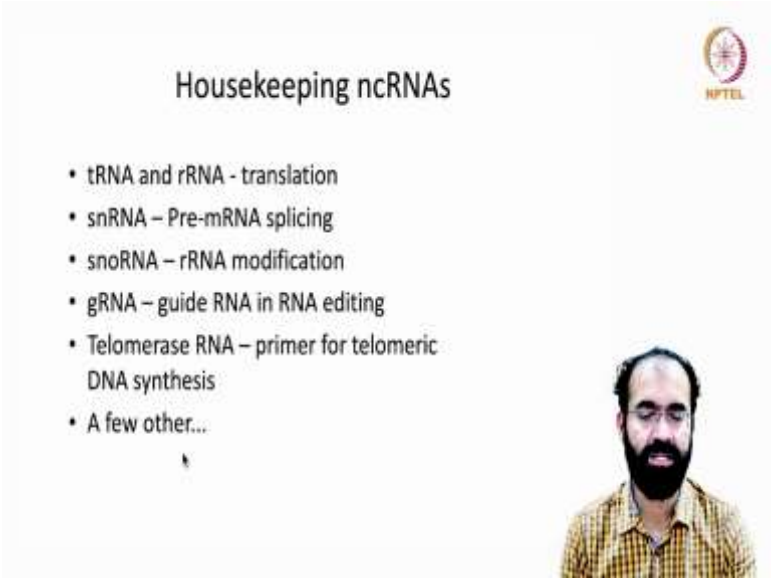
However, many genes their expression their regulation that is done via non-coding RNA that comes under regulatory non-coding RNA category. And this is a demand based or circumstance based. So, regulatory non-coding RNA can affect the expression of other genes at the level of transcription or translation. So, depending upon which is the necessary action to be taken.

So, regulatory non-coding RNA can influence the transcription of a gene that is whether or not a given mRNA should be expressed or not that is called transcriptional regulation. We you remember that we used two terms transcriptional gene regulation and post transcriptional gene regulation and we used the word with regard to silencing TGS and PTGS Transcriptional Gene Silencing, Post Transcriptional Gene Silencing.

So, non-coding RNA usually influence the transcription via the regulatory role it at the level of transcription. Some non-coding RNA such as micro RNA they influence the translatability of that mRNA in the cytoplasm. We know we yesterday's day for yesterday's or some of the earlier classes we discussed that how siRNA functions, how miRNA functions in blocking the forming the RNA induced silencing complex, how they are blocking the expression of mRNA.

Sometimes they cause the just translation translational arrest, sometime they cause degradation of those mRNA itself. So, in any case the gene regulation is via non-coding RNA either at the transcriptional level or at the post transcriptional level.

(Refer Slide Time: 15:17)



Let us see housekeeping non-coding RNA which are rather simple because by now we all know what are they. tRNA and ribosomal RNA useful in the translation protein production in a cell. snRNA important in the RNA splicing and snoRNA that is meant for ribosomal RNA modifications and gRNA, guide RNA in RNA editing.

The classic example is the CRISPR-Cas9 system where gRNA takes a small fragment of the RNA to the site where it is supposed to act which in turn will attract the Cas9 protein which will create a nick on the DNA. So, the gRNA was able to attract the Cas9 at a place where it paired with the DNA.

So, this is the simple I like last year's Noble Prize was given to this CRISPR-Cas9 system to Jennifer Doudna and Emmanuelle Charpentier. So, two women scientist received a Noble Prize for the discovery of CRISPR-Cas9. So, they studied a lot on the gRNA, guide RNA and the Cas9 system which is originally evolved as a defense mechanism in bacteria against viruses.

So, then comes the telomerase RNA they are basically for maintaining the length of the telomeres. We will see about telomeres more in detail in the future classes, but telomerase RNA is basically having a small template inside this protein which is telomerase. So, it acts as a primer for the telomeric DNA synthesis. So, it uses a small primer inside the telomerase enzyme and it makes a complementary strand DNA strand

of that primer that is in humans and lot of mammals it is T2AG3, TTAGGG this sequence will keep repeating.

So, that is the DNA ends and we call them as telomeres and that is one housekeeping function or housekeeping um. And remember telomerase is not expressed in every cell they are expressed only in germ cells and some young stage some and other situation is cancerous situation they are expressed.

Otherwise, they are not, they are not expressed and because of which the telomer length becomes shorter and shorter. But still, we can call it as a housekeeping gene. Because telomerase action depends on this small RNA which is acting as a template for the telomere repeats.

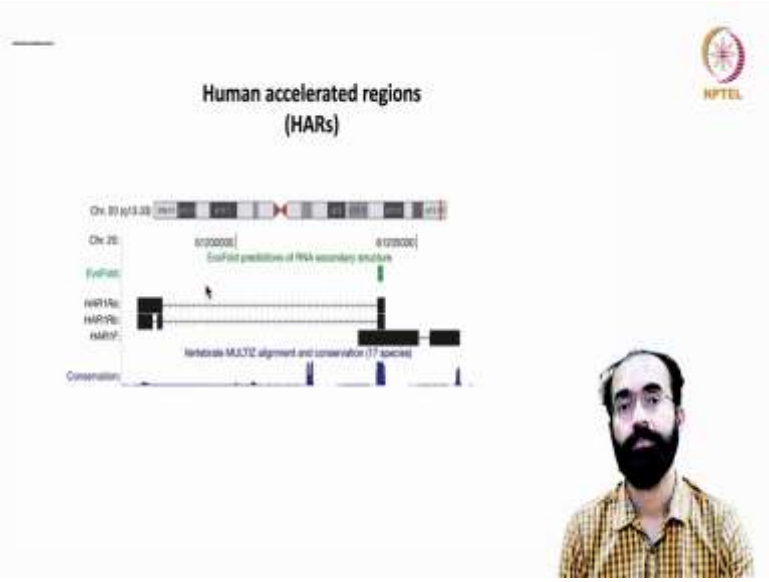
(Refer Slide Time: 18:18)

And few other examples are they are visible in the form of specific staining and they are speckles, paraspeckles, cajal bodies, germs, IGC's, PML bodies and nucleoli we all know at a place where the ribosomal RNA synthesis is taking place. And these area the speckles, paraspeckles etcetera you usually have snRNA and pre mRNA dominance. And they are also seen with specific staining in the nucleus of a given cell.

(Refer Slide Time: 18:51)

So, you can read this article for having a better understanding of how a non-coding RNA gene that can influence the expansion of human brain which has played a major role in the development of human assay intelligent species. So, an RNA gene expressed during cortical development and evolved rapidly in humans, this was published in nature as an article few years ago.

(Refer Slide Time: 19:30)

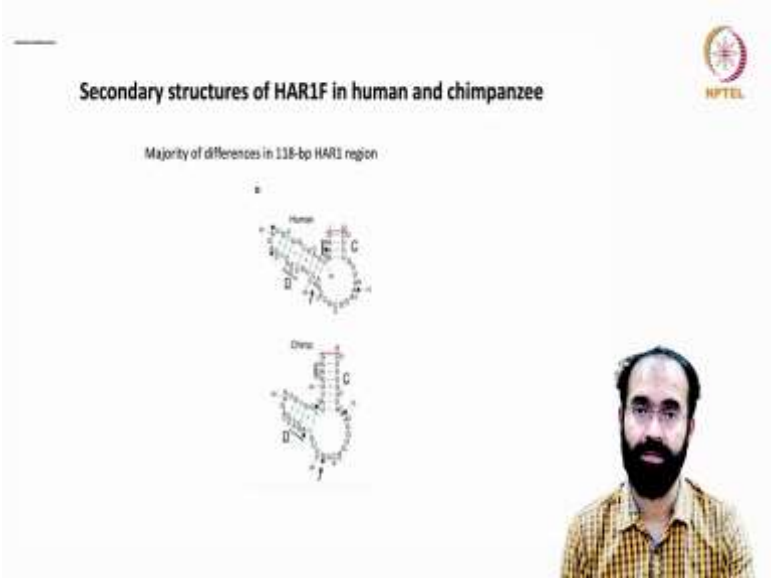


What is that gene? That gene is called Human Accelerated Regions that is the name of this gene and we short form we call it as HAR, HARs. So, it is situated in chromosome

20, we have got 23 pairs of chromosome. So, it is situated in one of the small chromosome, that is chromosome 20 is a small one and 21 is very important you may have heard about it 21 trisomy leads to Downs syndrome.

So, 21 is one of the small chromosome and 20 is little bigger than that. So, chromosome 20 has got this gene that is HAR1Ra, HAR1Rb and HAR1F. So, three genes its quite conserved gene and this gene is coding for RNA only not meant for producing protein. So, this RNA genes can be quite efficient in causing the cortical expansion.

(Refer Slide Time: 20:21)



So, secondary structures of HAR1F in human and chimpanzee are related. So, majority of the differences lie between human and chimpanzees between 118 base pair of the HAR region. HAR region within the secondary structure of this HAR RNA, this is human and this is chimpanzee. And you can see some base pair differences are there and this can influence the secondary structure of this RNA this variation can differences are in the 118-base pair at the HAR1 region.

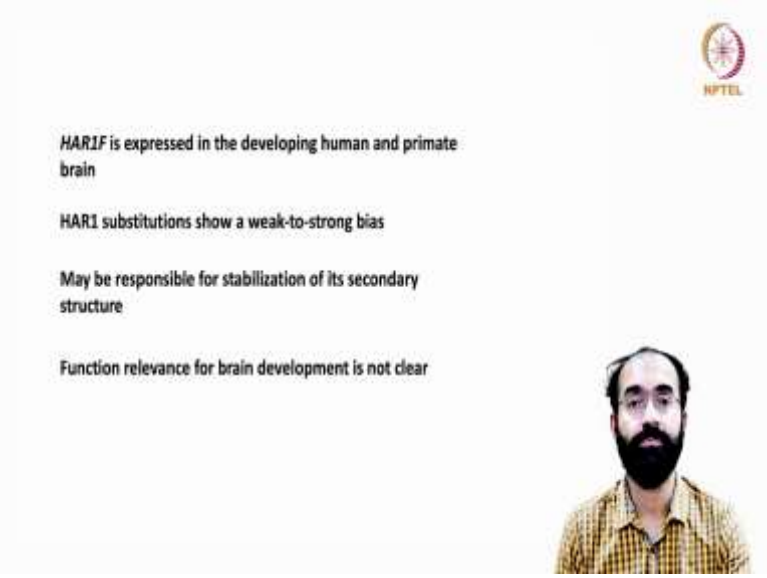
So, what is to be inferred from here is that the unique secondary structure we have in human HAR1, HAR1F here what is been discussed is capable of causing a expansion of the brain. That means our cerebral cortex constitute around 80 percent of your brain mass, 80 percent of your brain the credit goes to the cerebral cortex. None of the other animals have got such a huge cerebral cortex.

And cerebral cortex is supposed to be the major cause of enhanced cognitive functions in human like you can write a poem, you can appreciate beauty, you are able to communicate by talking etcetera. It is not mere size. So, do not think elephant has a bigger brain than human or blue whale will have much bigger brain than even an elephant.

So, that is not about the size, it is all about the communication and relative abundance, relative abundance of cortex is capable of doing lots of complex cognitive functions. And during developmental stage in the embryonic stage the HAR1 gene has expressed in very short time that can trigger. Because one of the tissues found very early during our development is the brain.

So, brain forms a very beginning of an organism development. So, cortex also has to happen during that time itself. So, HAR1F is expressed in the developing human and primate brain.

(Refer Slide Time: 22:40)



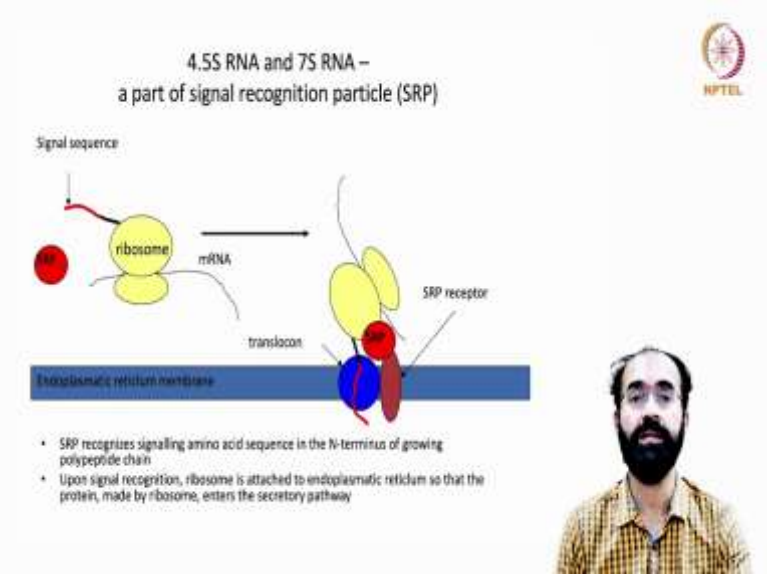
Primate means you know gorilla, chimpanzee, gibbon, orangutan, four primates are there. So, I think India has got only gibbon if I am not wrong and orangutan is seen in Singapore, Malaysia those kind of forests also and we do not have chimpanzees and gorilla which are mainly in the Africa.

So, total four apes are there means tailless monkeys. So, that is called normally we call them as the primate category. So, HAR1 substitutions show a weak to strong bias; that means, if we can replace in lower primates means is not so intelligent primates, you can see an expansion; that means, weak to strong bias; that means, you can increase the cortical size.

So, this is it is supposed to be because it can enhance the proliferation of this neurons. So, it is supposed to be responsible for stabilization of its secondary structure; that means, the HAR1 is capable of having a stable secondary structure unlike the counterparts seen in primates. So, functional relevance is attributed to the brain development, but the mechanism is not clear.

So, it is believed that HAR genes expression at a given time of embryonic development in primate embryogenesis humans have got an advantage compared to other primates which can cause a supreme brain structure and intelligence.

(Refer Slide Time: 24:21)



So, now let us see some importance of some of the RNA genes how they are contributing to the signal recognition particle. So, 4.5S RNA and 7S RNA part of the ribosomal complex. So, many a times the proteins can have a signal sequence and which is meant for secreted out of the cell. If a given protein has to function outside the cell it will have some sequence especially in the N-terminus of the protein that will be recognized, but who is there going to recognize? That is recognized by signal recognition particle.

And this signal recognition particle is nothing but a complex of RNA. And they are as soon as they are produced that will be recognized and that will tether the whole RNA, ribosome and the newly formed protein complex to the endoplasmic reticulum membrane.

Endoplasmic reticulum is an organelle that has got a connection to the outside of the cell. You may have studied in school days smooth endoplasmic reticulum and rough endoplasmic reticulum. Rough endoplasmic reticulum are those endoplasmic reticulum that contains ribosomes attached onto them and while those proteins are supposed to be entering into the lumen of the ER which eventually find its way outside the cell.

Some of them will pass through the golgi complex where it has to do post translational modification and eventually it is going out of this like hormones, many enzymes etcetera. Like enzymes are after eating food, enzymes are secreted out of the cells that is why it is able to act in your intestine or your stomach.

So, many of them are secreted out. So, signal recognition particle recognize signaling amino acid sequence in the N-terminus of the growing polypeptide chain. Upon signal recognition the ribosome is attached to the endoplasmic reticulum so that the protein made by ribosome enters the secretory pathway.

So, that the protein will enter into the lumen and it can eventually find its way out. So, signal recognition particle discovery has fetched Noble Prize. It is so important and so exciting discovery. So, we will continue to learn about non-coding RNAs especially signal recognition particle in the next class.

Thank you.