Course Name: I Think Biology

Professor Name: Dr. Sravanti Uppaluri

Department Name: Biology

Institute Name: Azim Premji University

Week:3

Lecture:16

W3L16_Non-Coding RNA

Hello, and welcome to our lecture on non-coding RNA as part of the iThink Biology NPTEL course. So in the last class, we discussed gene regulation, right? We talked about how nearly 2% of our DNA codes for protein. This was something that was discovered when the Human Genome Project was being conducted. And that means that of the several billion base pairs of DNA that we have, only 2% of it is actually transcribed into mRNA, and this mRNA, in turn, is translated into a polypeptide or a protein.

So what happens with the remaining DNA? Why do we have so much DNA that actually doesn't encode for protein? It turns out that even though most of it doesn't encode for protein, nearly 70% of our DNA is actually transcribed. So it's made into RNA. And we call this RNA that is does not code for a polypeptide or a protein. We call it non-coding RNA.

So this is a picture that we saw in the last class. And it was really a modification of what we know to be the central dogma. And the central dogma is really normally characterized as being this very linear thing where DNA encodes RNA. And so DNA is transcribed. So we go through the process of transcription to make RNA.

So a subset of the DNA sequence is transcribed into RNA. And the RNA can then be translated into a protein. The transcription process happens inside the nucleus. And the translation process happens in the cytoplasm. Now we heard a little bit about how the same RNA molecule through alternative splicing can actually give rise to different isoforms of a protein.

We also talked about how some RNAs don't make protein at all, but they make non-coding RNAs. And what is the role of these non-coding RNAs? It turns out that they actually have many roles and many functions in inside the cell. And when the central dogma was first proposed, I

don't think anyone ever imagined the sort of diversity in functions that we now know of for non-coding RNAs. So let's talk a little bit more about, what non-coding RNAs are?

So you already know about tRNA and rRNA. So tRNA is found in the cytoplasm, right. So let's make notes as we talk through this. So it's found in the cytoplasm. So is rRNA, right, and both of these are part of the translational machinery, right.

rRNA is required for the, because it has the catalytic site that is required for translation and t in the, sorry, as part of the ribosomes. And tRNA is required because it actually recruits the correct polypeptide to the mRNA in order for translation to occur, right? Now, the remaining RNAs that we are talking about are really the ones that are of most interest to us in this lecture because they are the ones that are sort of part of the gene regulatory function, right? So let's start with small nuclear RNAs. So these, as I said, are, as a name implies, they are found in the nucleus and they are usually around 150 nucleotides in length.

And what they do when they, as they are found located in the nucleus, they actually regulate transcription factors. So we discussed in the last class that transcription factors are proteins that help regulate gene expression. So they help decide whether a protein should be expressed, should be turned on or off, right? So small nuclear RNAs-snRNAs bind to transcription factors and can help a transcription factor bind to its target site or prevent it from binding to our target site, right?

Okay, MicroRNAs are generally shorter in length. They're around 20 to 24 nucleotides in length. And what they actually do is that they are, so they're abbreviated as miRNAs. You will see this everywhere. And they are actually single stranded RNAs, right? They are, in that form, hairpins once they are transcribed.

So what do I mean by hairpins? So imagine that you have a sequence of RNA with different bases. So in the case of RNA, it's A, U, Gs and Cs. Right? And this same strand might fold in on itself, where the bases now become complementary. Right? So this is what is called a hairpin. And microRNAs are usually involved in post-transcriptional regulation. What do I mean by that? So once transcription has occurred for any kind of gene, microRNAs can help sort of prevent the expression of that gene if it's required. So it can block essentially the mRNA of a protein that's, that the cell is trying to express. OK, long non-coding RNAs tend to be over 200 nucleotides in length. That's why they're called long non-coding RNAs. And they actually can, they have several roles.

They occur in order to regulate transcription like mRNA, miRNAs. But they also play a role in regulating miRNAs and siRNA, which you will hear about next. They may also, long non-coding RNAs may also help in chromatin modification. What do I mean by this?

So if you wish, if you, if a cell wants to express a gene, the chromatin first has to be decondensed and unwound from the histone proteins. That's what we talked about in the last class. And long non-coding RNAs can help facilitate this process.

Okay, small interfering RNAs are similar to microRNAs in that they are about the same length, right. They are 20 to 24 nucleotides in length. But they are not endogenously expressed. What do I mean by this?

So microRNAs usually have a microRNA, a corresponding microRNA, DNA sequence within the genome of the organism. And that is transcribed to form single-stranded RNA, right? To, And this single-stranded RNA can then go on to be part of the post-transcriptional regulatory machinery. Small interfering RNAs, on the other hand, are exogenous. So they may come in, they may be sourced from a virus, or they may come in from the external environment. But the presence of these small interfering RNAs is also part of the post-transcriptional regulatory machinery.

Okay. And because small interfering RNAs can be exogenous, they also have biotech applications, biotechnological applications. And what I mean by that is, imagine you can interfere with gene expression or control gene expression with the use of siRNAs. Because they are exogenously induced anyways, we can also do it in the lab, right? And so we can control gene expression using the siRNA machinery. So in today's class, we will focus on microRNAs and small interfering RNAs, okay? Small interfering RNAs, siRNA, is often also referred to as RNAi, which is RNA interference.

OK, so I'll start off by illustrating how microRNAs were discovered in the first place. They were discovered in C. elegans. C. elegans is Caenorhabditis elegans. It's a round worm that's used for research. So the life cycle is illustrated here. It starts off as an embryo. This embryo hatches and forms an, a worm.

And that worm goes through four larval stages, right? So that's what you see here, L1, L2, L3, L4. Finally, it reaches adulthood, where it is reproductively mature and then can lay embryos and so on. So that's the life cycle of C. elegans. And you can imagine that to go through every larval stage, there is a mold. And in order to induce that mold and to pass through these different larval stages, there has to be some sort of gene regulation that occurs, right. And it turns out that in the early 2000s, it was discovered that there was, there was a mutant that actually never went from the L1 to the L2 mold, right? And so basically, the gene expression patterns in some ways continued to stay what they should be in the L1 stage. And they were able to identify that this mutant occurred or this, this mutation had the phenotype because of a particular gene called lin-4.

Yeah. And it turned out that lin-4 actually encodes a microRNA. Okay. So it doesn't actually ever get translated into a protein. Rather, it forms a single standard RNA. And this single

standard RNA actually prevents the expression of a protein called lin-14. And it does this by helping the degradation or rather preventing the translation of the lin-14 mRNA into the lin-14 protein. Several years later, something similar was found for the transition from L4 to adult. And in this case, it was an mRNA, miRNA called let-7, which prevented the expression of lin-41.

This is a blocking. Yeah. Right. And so how does, how does this repression occur through miRNAs? So if you look at expression levels, so you, if you recall the example of lin-4 and lin-14, so lin-4 is the miRNA that I talked about, right? And as lin-4 starts to get expressed when the worm crosses the L1 stage and needs to go into the subsequent larval stages, what happens is that this lin-4 actually causes lin-14. This is the mi. So, in blue are the mRNAs, the levels of RNA expression.

In black is the level of the protein expression. So because lin-4 is being expressed in larger amounts, it prevents lin-14 from being translated into protein. And, so lin-14 protein levels also start to drop. And that is the wild type phenotype. If lin-4 is removed, so if there is a mutation in lin-4 and it can no longer repress lin-14, then the transition from the L1 larval stage to the L2 and L3 and so on are actually, is actually prevented.

So that was the sort of background with which miRNAs were first discovered. So people never imagined that it was actually an RNA molecule. When they traced the mutation back to lin-4, they assumed that lin-4 encoded another protein. But instead, they found that lin-4 was actually for the the lin-4 gene, I guess, or the sequence actually encoded a micro RNA. Okay, and so again, we, we like to look at these expression levels in the cyclical pattern because it helps us really understand how, what the relationship between different either protein or RNA molecules is, right?

So in green, you see that between the L1 and the L2 stage, the lin-4 RNA, right, so this is a microRNA, expression really increases to the point that lin-14 expression drops significantly by the time almost to none to zero by the time it reaches the L2 stage. Right? And what this means is that, it allows for the transition to the L2 larval stage. The same is the example for LET-7, it starts getting expressed, LET-7 again is the miRNA. So let me specify that here just to prevent any confusion.

So the miRNA, this is also a miRNA. And for the L4 to adult transition, the let-7 has to be expressed. This prevents LIN-41 expression. And this LIN-41 is a protein that is normally expressed. But because its mRNA is is prevented from being translated further into L LIN-41 protein, the protein starts to drop. And so you can see that there's really these very strong sort of switches, right? That's what these miRNAs do.

So it's not a very gradual switching off, but rather it's a strong switch. And that really, that really

is required when you have very clear events in development, right? So there is no in-between stage between L1 and L2. You won't or you don't. And so you reach the L2 stage or you don't at all. So these kinds of switches in levels are really important when you need, when you need a clear transition from one stage to the next.

Okay, so let's try to understand what microRNAs do. So this looks like a complicated slide, but it's actually not. So we'll start off at the very top. So this is the miRNA gene, right? So, this is the DNA and on the DNA is the miRNA gene. This gets transcribed as is shown here. And once it's transcribed, it automatically forms a hairpin.

And that hairpin of course depends on the sequence. So that hairpin will form based on the complementarity between bases. So for example, the lin-4 miRNA that I talked about earlier is shown here, and you can see that it's folding as a hairpin because of complementarity between these sequences here, right? You can also see if you look very carefully that if you were to shift this hairpin a little bit, so it's folded like this, but if you were to shift a little bit, it's also a possibility for the configuration of the hairpin. This becomes important a little bit later. So what you see is that once the miRNA is formed and it has formed this hairpin, so it's folded on itself, it will get processed and then shifted out from the nucleus to the cytoplasm.

When this happens, Dicer, which is again a protein complex that is very important for gene regulation, will bind to this hairpin. Together with this, the Dicer complex actually converts the hairpin into a into single-stranded RNA, right? So you can see now that the RNA has been cut up right, and once a single-stranded RNA complex has been formed, there is another complex that forms around it called the RISC complex. And this is again a bunch of proteins that form around this, these two strands of RNA that are bound together. Now the RISC complex facilitates the binding of a single strand.

Now remember where this strand has come from, right? It's come from the original hairpin that was transcribed from the miRNA gene. And this single-stranded RNA, as it's in the RISC complex, will bind to the gene of interest. And once it does this, this binding can repress the translation of the of the mRNA, right. So this is an mRNA of interest. So if you think about lin-4, which is what I mentioned before, lin the lin-4 miRNA would bind in this case the lin-14 mRNA.

And once this is bound, translation is repressed in several ways. And these are listed here. Right. So this is, repression of translation. And this happens either by preventing the initiation of translation at all, so the ribosome can't move forward. It can also cause the cap from the mRNA to fall off, which means again translation does not occur.

The binding of the mRNA, miRNA can also block elongation of the polypeptide chain. And it

can also degrade or deadenylate, so prevent this poly-A tail from forming. And together, all of these things mean that, the translation of this mRNA molecule into protein is prevented. So this is a really important sort of pathway.

And a few things that I would want you to remember. First is the Dicer complex that turns the pseudo double-stranded RNA, or rather the hairpin RNA, into single-stranded RNA. The RISC complex, which helps facilitate the binding of the mRNA to the mRNA of interest. And then the different ways through which translation can be repressed. Okay. So it turns out that this complex of Dicer is important not just to mi miRNA, not just to microRNAs, but also to small interfering RNAs.

And how this happens is that the hairpin is the miRNA. But as I mentioned earlier, siRNAs are double-stranded RNAs that are exogenously that are found inside the cell through exogenous sources. So the Dicer complex actually turns these double-stranded RNAs into single strands. And that is really important for the eventual silencing of mi of mRNA from a gene of interest. Right? Okay. So moving on to why, moving on to small interfering RNAs, why are they interesting at all? So they are not so different in the sense that, different from miRNAs in the sense that they help prevent gene expression or gene, sorry, mRNA, mRNA translation.

But siRNA or RNA interference can be used also as an exogenous source. So again, this tech, RNA interference was developed as a technology by Craig Mello and Andrew Fire, and they received the Nobel Prize for this. And they received the prize for this because of the fact that RNA interference can be used to targetedly repress a gene expression. Right? And what they did was they showed that you could take double-stranded RNA, right? And that's what you see here. And this RNA would have to be complementary to a specific gene that you wanted to suppress.

And if you injected a parent before the parent produced, this again is done in C. elegans, before the parent produced any progeny, the progeny would then take up this double-stranded RNA. And through the Dicer and RISC complex that we discussed, specific genes would be targeted. And in this case, they targeted muscle genes, right muscle a muscle protein, and this resulted in twitching. And an important part of this discovery was that in order to conduct RNA interference, in order for RNA interference to work, you really had to put in the double-stranded RNA.

You could neither use the sense RNA or the antisense RNA. So by themselves, the single-stranded RNA is not recognized by the Dicer complex. So you need the double-stranded RNA in order for RNA interference to work in the first place. Right? And why this is this whole this idea of RNA interference became very important, first because we were able to understand that RNA could be used as, as a regulatory, as part of the regulatory machinery for gene expression. But RNAi could be used for many biotechnological applications. Right? So it could

be used in the lab to targetedly suppress a specific gene, right, depending on the sequence of the RNA that you injected into the organism.

You could target a specific gene and shut it down or turn down the amount that was being expressed. So this was a really big breakthrough, right? Because once you turn down a specific gene, you can see what the effects are. And once you see what the effects are, you can infer what the function of a gene is. Right? So this is a very important way in which we can understand genes and their function. Not only that, because you can use RNA interference to specifically suppress a gene, it may also be useful for disease treatment.

Okay, so we've already covered a little bit about how this knowledge helps, right? So let's try to summarize it. First, understand more about gene expression regulation. We can use it for, let's say, genetic studies for gene function. We can also target specific genes, and this may help in disease treatment. Right? And we know, of course, that there are many diseases that arise from, say, single base pair mutations, and perhaps targeting these proteins might help, or that overexpression of specific genes can lead to disease.

So these kinds of things could actually help. And it turns out from 2006, when Mello and Fire received their Nobel Prize, there actually are several therapeutic drugs now, or therapeutics, they're called RNAi therapeutics, that are available, especially for metabolic disorders, for using RNAi technology to help treat metabolic disorders. So that's a really interesting outcome of these studies.