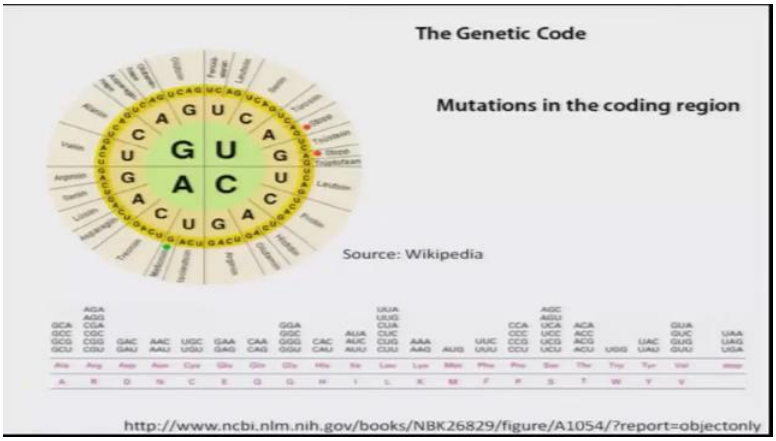
Human Molecular Genetics
Prof. S. Ganesh
Department of Biological Sciences and Bioengineering
Indian Institute of Technology, Kanpur

Module - 01
Lecture - 03
Fundamentals of Central Dogma – Part III
(DNA, RNA and proteins; mutations)

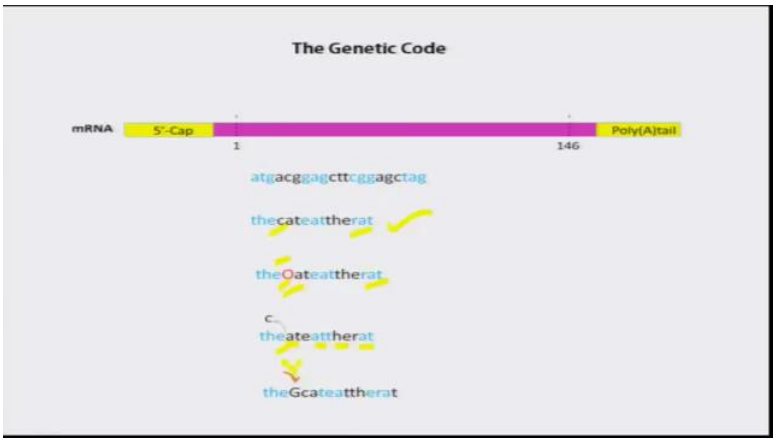
Welcome back to the third lecture of this course Human Molecular Genetics.

(Refer Slide Time: 00:17)



So, what we discussed in the previous lecture is that how the changes in the DNA can alter your survival, whether it makes you fit or unfit. So, today we are going to look into or in this particular lecture we are going to look into how mutations really affect the protein, when the mutations fall in the coding sequence. So this, what is shown here in this slide is, that, is the genetic code. So, what are the different codes that give the signal for a particular amino acid to be added in the given region of the protein?

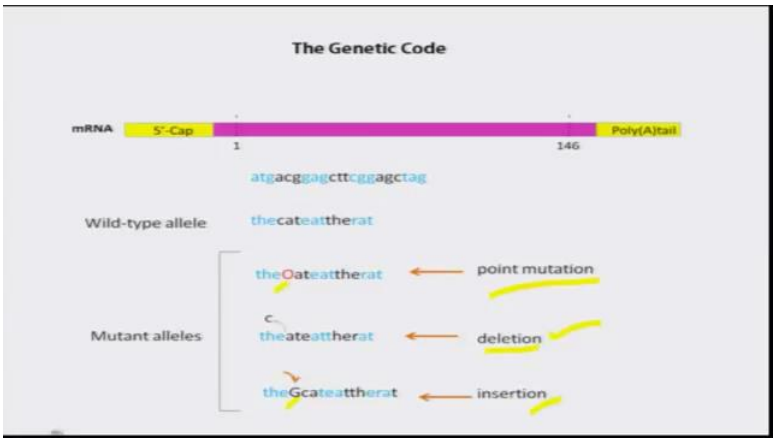
(Refer Slide Time: 00:55)



It all begins with the RNA, which has got the sequence and the sequence is what is called as translated. So, this is something we discussed that each code is a triplet and each triplet give an information, as to which amino acid should be added. So, let us look into the example that we discussed before, the English alphabets which gives a meaning in a sentence. Here the words are colour coded. Each word is three letters; 'the cat eat the rat'. Now, what would happen if this was the genetic code and if there are changes here as to how the meaning of this particular sentence is altered, we will consider that as an example, then, we will get back into the genetic code per say.

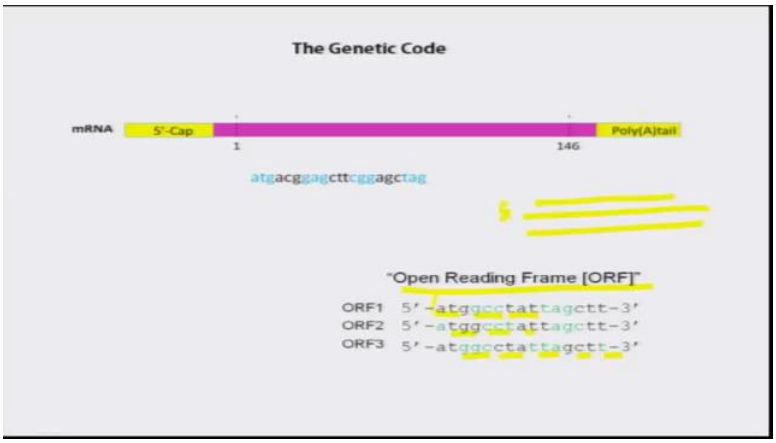
Let us see; so, there are three different changes that I have shown here. The first one is the letter C is transferred to O. Now, you do have all the words that are three letters, but the meaning of the sentence is altered. In the first what it gives the meaning is the predator cat is able to consume its prey which is rat. Here, because of one particular letter being changed in this particular word that is cat, the meaning of the sentence is misleading. Now, the oat became the predator for rat, which is not possible. So, it does not convey a meaning. It is because of a change, change or substitution, one letter is substituted by the other. It could also have a condition, wherein in the sentence by mistake you have not typed a letter. Here for example, the C was not typed; as a result you have converted each three letter as a word. Now, these three letters really doesn't make any sense or you could have typed an extra letter in a place where it is not required or should not be there.

(Refer Slide Time: 3:10)



As a result, again if you convert three letter words, it does not convey any meaning and this is what you call as mutations and these new forms of the sequence are called as mutant alleles and the one that make sense is called as wild type. So in other words, for the same gene it could have hundreds of alleles that are possible, not necessarily in the same individual but in the population, like what you see. But if it affects the function of the protein, you may not see that in the population. Now, what does it really mean? All these conditions wherein one particular base is replaced by other we call them as point mutation, because they are restricted to a particular base. But, if a base is removed like you see here, you call that as deletion, meaning deletion of a base or you have inserted a base, which, should not be there; that you call as insertion. So, these are the different nomenclatures for the mutations that you see.

(Refer Slide Time: 4:13)

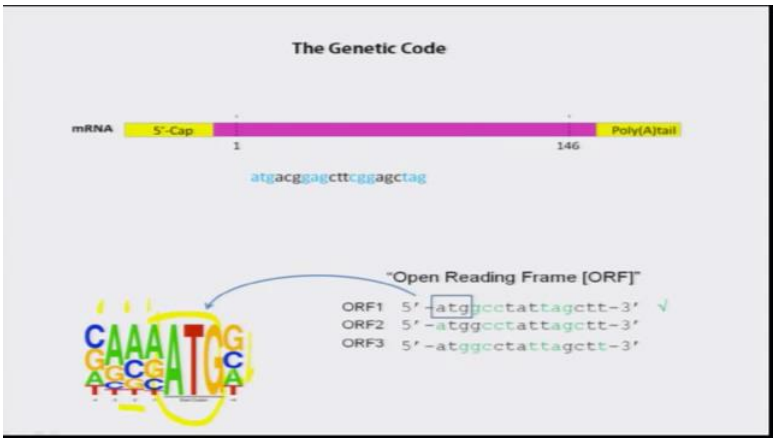


Let us look into the examples. Now, what we discussed is that a sentence should give a meaning. For that you make three letter words and each word gives you a meaning. Now,

how do you know where to begin? For example, if you look into a double stranded DNA, the strand that gives the coding sequence is normally called as sense strand and that is what gives you the message as to what amino acids are added, in which sequence? Therefore, if you look into a DNA sequence, which is normally shown 5' to 3', you can really predict three different frames what you call as open reading frames or ORF. You can start with A; therefore you have ATG, GCC, TAT and so on, or you can start position 2, then you have altogether different sequence or you can start from the third position, then it gives different combinations. It all depends, which frame you are beginning, you are making the construct or constructing a word. So, that is what is called as open reading frame.

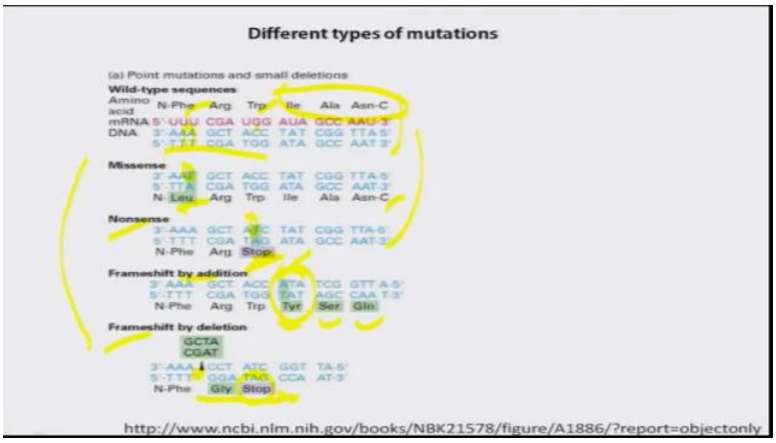
So, in most of the RNA that you see here, you have only one functional open reading frame, meaning you have to start making a sense out of the sequence, from a particular point in the sequence. So, how the cell knows?

(Refer Slide Time: 5:44)



Invariably, it starts at this sequence. It starts with ATG which itself gives you a message for making methionine in that particular place or the first space of the peptide. But, it is the surrounding region, the upstream sequence, for example here, majority you have AAA and followed by ATG, followed by GC or A in that frequency they are present. Normally, you have such kind of sequences, which is called as Kozak transitional, initiation sequences that is considered as the first word for decoding the message that is there in the RNA. So, if that is fixed for example, like what you see here, then the rest can easily be decoded and they give the meaning for different amino acids.

(Refer Slide Time: 6:44)



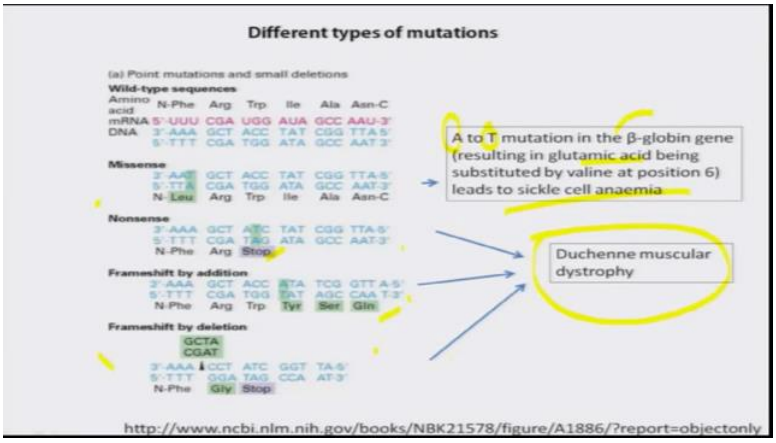
So, this is how cell is able to identify the sequences that are present in the RNA and translate them into the protein. So, these are examples given to explain how for example point mutation, one base change, can affect the way the protein functions and you can see here, this is wild type meaning present in majority of the individuals and these are the different mutant forms or mutant alleles. So, you have a base here for example the A is converted into T. As a result, now it gives a different meaning. Here you have an amino acid that is converted into a different amino acid. So, this is called missense, meaning they did not give a correct sense as to which amino acid should be present in that particular peptide. So, this is called as missense mutation, because it depends on how that change in the DNA has affected the protein. So therefore, here it has affected the peptide by replacing one amino acid with the other, therefore it is called as missense mutation.

But, it could also have situation, wherein a change in the base would introduce what is called as stop codon in place of an amino acid. So, this stop codon, as a result the proteins will not go any further, it will be terminated there; synthesis will be terminated and that is called as nonsense, meaning it doesn't allow any new amino acid to be incorporated in that position. Now, these two mutations are called as point mutations, because they change the base. For example, here A to T, here C to T. But, you could have other conditions like we discussed using cat, rat model, like deletions.

So, that is something that is shown below; for example, here you have the same sequence. Now you have the three amino acids that are coded like what you see here, but over here what you had was a new, know, base is inserted. As a result, know, if you make the triplets,

then you are going to find a new, know, amino acids being added and they are not what you see here and this is happening because of, there is a shift in the open reading frame because of the insertion of a new element and this is called as frame shift mutation. It also, the frame shift can also happen because of, I, deletion, like for example here, you have lost a few sequence. As a result, the reading frame is shifted and that gives you a stop codon. Therefore, the peptide no longer can be made; gives the true example of the sentence that we, you know, just now discussed about cat and rat?

(Refer Slide Time: 9:59)

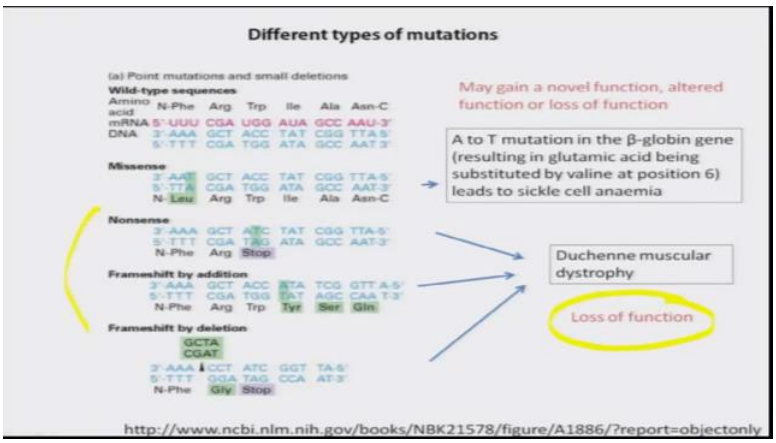


So, let us see how really they can cause disease. For example, one of the common text book examples for a missense mutation, meaning one amino acid being replaced by another because of a change in the DNA is sickle cell anaemia, wherein for example a mutation from A to T in the gene that codes for beta globin, results in the formation of a mutant protein globin. As a result, it affects the physiology and the cell that look very, very abnormal and that is called as sickle cell and also it is associated with anaemic condition. So, that is why that is called as sickle cell anaemia. This is very classic example; you can read more about it. So, this change A to T brings about change in the amino acid from glutamic acid to valine, at position 6 of this particular beta globin gene.

Now here you change the amino acid, therefore it is a missense mutation. Now in other condition, wherein you have, know, stop codon or nonsense mutation wherein you have introduced a stop codon or there is a frameshift, where reading frame has been altered, because of insertion or deletion, can affect the gene function. One example is one form of muscular atrophy, wherein the muscle degenerate, give up. Therefore you loose all your

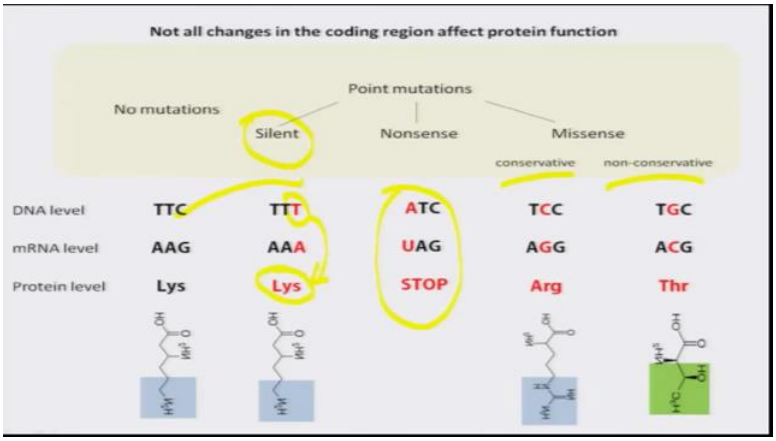
strength resulting from a particular gene which codes for a protein which is present in the muscle. Now, here each one of the mutation that you talk about, nonsense where there is a premature addition of stop codon, peptide is not being made or there is a frame shift either because of deletion or insertion, all of them likely to affect the protein function or even its synthesis, because it will be a very small peptide, it cannot even function.

(Refer Slide Time: 12:05)



So, they result in what is called as loss of function. The protein, loses its original function; it is almost like not having the gene. This happens because of these three different kinds of mutations. But, missense mutation may or may not result in loss of function. There are examples, wherein an amino acid substitution resulted in the loss of protein function. There are also many examples, wherein a change in the amino acid resulted in the protein acquiring a new function. The new function could be anything from being toxic to the cell, becomes insoluble, kills the cell or it could be for example it's an enzyme. A change in the amino acid makes the enzyme more active, which may not be desirable for the cell or any function that the particular amino acid modulates in that protein is altered. So, it acquires something pretty novel, which is not desired. So, these are two classes of mutation depending on how they affect the gene function. So, there are a number of examples for explaining the loss of function or gain of function some of which we will be discussing later.

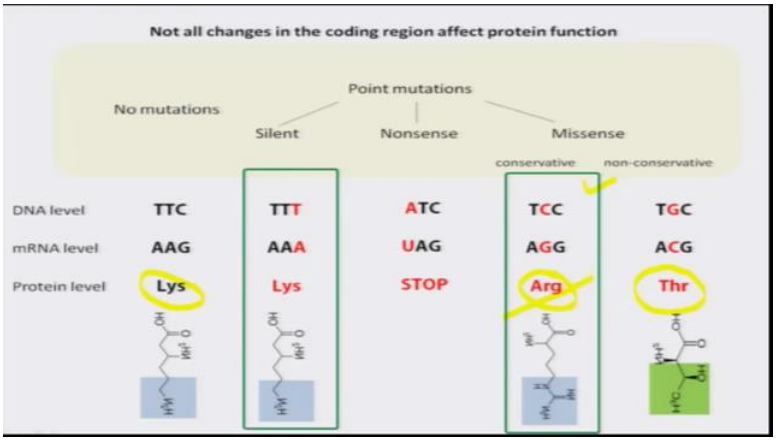
(Refer Slide Time: 13:32)



So, let us look into that. Is it that all the changes that are present in the coding regions affect the protein function? The answer is no. It is not necessary that every change that is present in the coding region should affect the protein function. Let us see some of the examples. There are mutations which you call as silent mutation. What does it mean? You do have a change in the base, for example from C to T, but it did not alter which amino acid should be added in that position. As you have seen, there are amino acids which can be coded by more than 4 or about 6 different codons. So, even if there is a change still, that code gives the same meaning, as to which amino acid should be added. So, even if there is a change in the DNA, it would not affect the way the protein is being synthesized. It would still have the same amino acid in that place. So, such mutations are called as silent mutation, meaning it does not really affect the way the protein is being made.

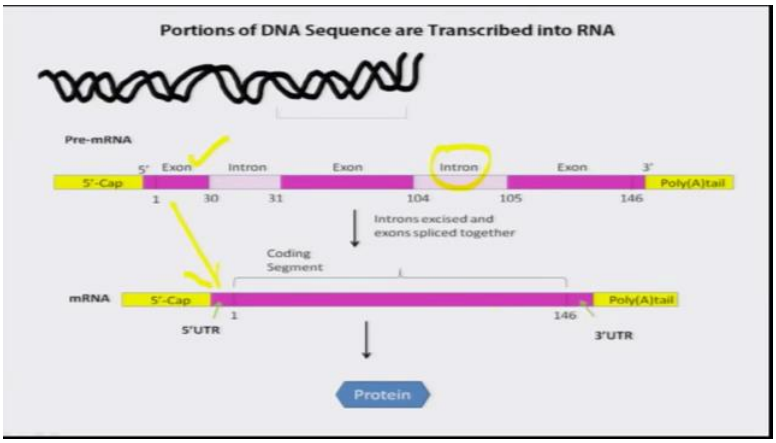
For example, nonsense mutation, invariably it would affect, because it introduces a stop codon. Therefore, the peptide is no longer a full peptide. It would affect its function, but you do have conditions wherein you have two distinct effects, missense mutation having two different effects on the protein, which is called, one is called as conservative, other one is non-conservative. What does it mean?

(Refer Slide Time: 15:03)



Let us look in here. For example this conservative change, wherein you have arginine that has come in place of lysine, but still allow the protein to function the way it does. There are many places in the population, wherein you have individuals having different amino acid in the position; like for example lysine to arginine, still that protein is able to function without much of a compromise. But there are changes; for example here if lysine is replaced by another amino acid, for example threonine, it may affect the way the protein functions. So, such changes wherein the change in the amino acid does not affect the protein to that extent, because it is very similar amino acid being replaced, are called as conservative substitution and these are normally present in the population. If I sequence 50 individuals in the class and I would find that at least 4, 5 of them would have amino acid like what you have seen here, lysine to arginine and they are normal? So, these are called as conservative substitution.

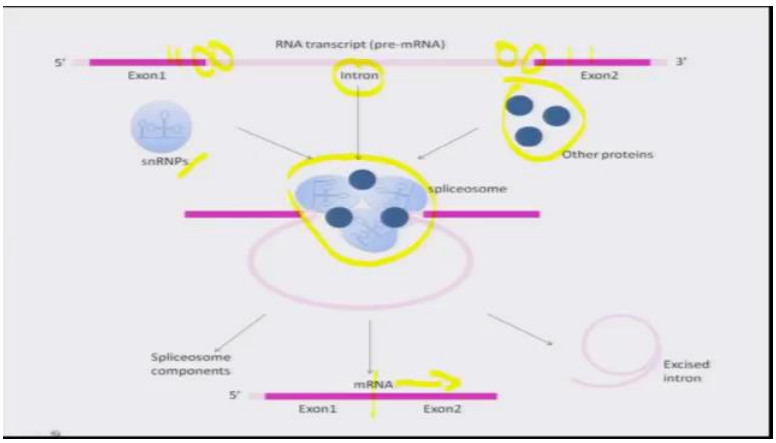
(Refer Slide Time: 16:09)



That is about how mutations in the coding region could affect the way the protein functions. Now, we are going to go and look at a different group of mutation, which need not be present

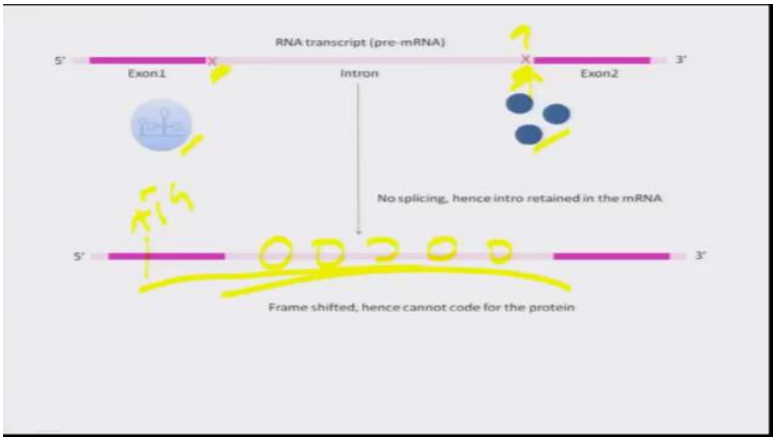
in the coding sequence, but still may affect the gene, the way they function. So, what we discussed was that you have genes, which have got exons which are retained here in the matured RNA and regions called introns, which are spliced out during the splicing event which are not present in the matured RNA and then the matured RNA is translated into protein.

(Refer Slide Time: 16:53)



This process of splicing, removing the intronic region is a very complex process. It involves many proteins. For example there are RNP's which is a complex of protein and small RNA's and there are also many other proteins which help in cleaving or cutting the RNA and joining the RNA. They together form what is called as the complex, which you call as spliceosome, which assemble on exactly the boundaries of exon and intron and then remove the intron and join the two exons together and this has to be a precise mechanism. Otherwise you may, even if you add one base more or one base less, then any sequence that is present after the first exon will not be read properly, because it is going to shift the reading frame, it is going to alter the reading frame. So, it is such a precise process and it happens for every copy of the mRNA that is being made in your cell; so, it is such a precise mechanism and how does it really work? They identify sequences that are present even in the intronic region and there are many elements even present in the exon, which bind and they know exactly where they have to cut the RNA and join together. So, it a very elaborate mechanism; we will not get into the details.

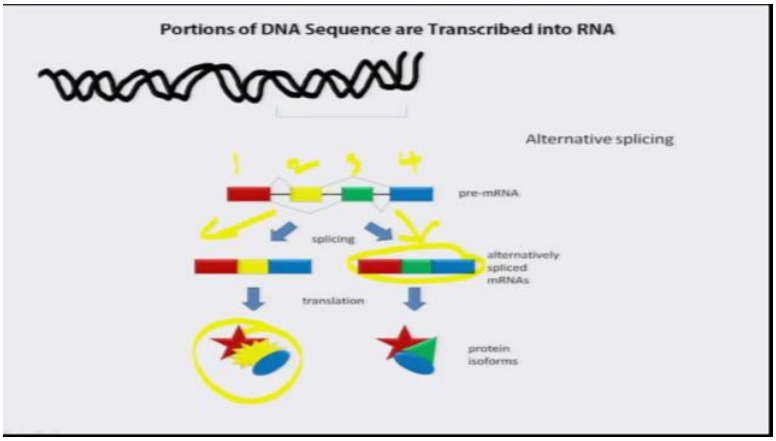
(Refer Slide Time: 18:19)



What I would like to highlight is that there are sequences even that are present in the intronic region, but these sequence are very, very critical for these proteins, which help in cleaving or cutting the RNA. If these sequences are altered for example, T should be there in this place and the T is replaced by G, for example. Now, these proteins will not be able to recognize them. So, as a result, you would have an mRNA in which the intron is retained, not deleted. Now, if the intron is retained, then you have for example ATG here, now what would happen? The translation machinery is going to read continuous and here, it is not going to give you the message that is required for making the protein. As a result, you have lost the protein; you are unable to make the protein that is required by the cell.

So just like the way we have looked at, sequences that are altered, changed in the coding sequence may affect the protein; similar way, there are critical sequence that are present in intronic region, if they are altered, RNA will not be spliced; as a result, it will not make the desired protein. The coding region will be altered or the reading frame will be altered. So, that is again to emphasize that even in intronic region there are elements, there are sequence that are necessary for the normal splicing process and mutations may affect the splicing process.

(Refer Slide Time: 19:56)



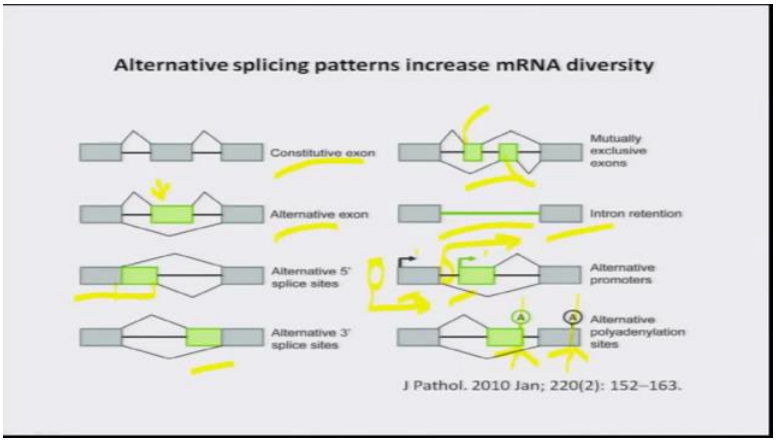
Now, we are coming to a new concept, which is called as alternate splicing. What they are? What we have looked at in the previous example is that a gene has got multiple exons and the exons are joined together by a process called as splicing and then you have a matur transcript, which are read by the translation machinery and you have the peptide. Now, what happens in some of the genes is that not all the exons are joined together always. For example, we have given here two examples. In example 1 you have, exon 1, 2 and 4 are joined together to form a particular protein. In example 2, you have exon 1, 3 and 4 joined together and then, that gives you a message for a particular protein. So, the same gene and the same transcript, depending on how different exons are joined together, gives two different signal as to what kind of peptide it has to make.

Now, this is very interesting and there are a large number of examples in the human genome as to how many such genes exist. In fact, if you compare different organisms, humans really do not have a large number of genes as compared to so called lower organisms. But, perhaps the diversity comes in the way we are able to make several proteins out of the same genes, but by using differential splicing, what is called as alternate splicing. There are examples; for example, in drosophila, there is a gene that can make up to 30,000 different proteins, all because of alternate splicing of the RNA of one particular gene. So, the diversity is enormous. We have not understood all these processes, but what we know is the alternative splicing is very, very critical and you look into this example.

For example here the sequence that are involved in creating or generating this particular transcript, wherein exon 2 is not retained may be very different as compared to the sequence that are being used by this particular transcript. So you may have a mutation that may affect

one of the transcripts, not the other and depending on the mutation you may have a different phenotype. These are all possible.

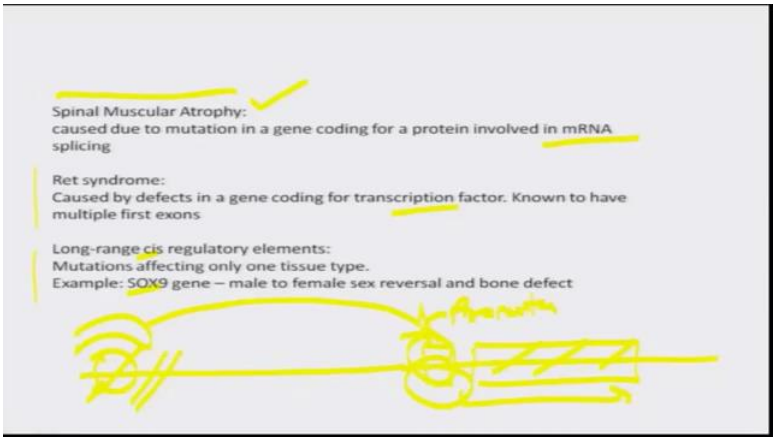
(Refer Slide Time: 22:34)



Again, this is an example to give how complex the alternative splicing is in our genome. For example, you call this as a constitutive exon, meaning these exons are present invariably in every transcript that is being made. You call here as alternative exon, because this may or may not be present in every transcript. Here you have an exon which can be full or partial, depending on where the splicing took place; likewise here and these are called as mutually exclusive, meaning if this exon is present this would not be and if this exon is present this would not be and you have genes in which all of a sudden a particular transcript may retain the introns. As a result, the coding frame is shifted and you may have two different first exons and this also brings in challenge, because, the cell needs to have two different promoters.

This may be operated by a different promoter and this may be operated by a different promoter or you could have two different last exons depending on where the translation or transcription is terminated. So, these are the complexities and you can imagine how a change in the DNA sequence can affect a particular isoform, which you call as a variant of the transcript and how that could affect the cell function in an individual.

(Refer Slide Time: 24:10)



So, it is extremely complex and this complexity is also highlighted in these findings. For example you have a condition called spinal muscular atrophy, wherein your muscle tissues degenerate. As a result, you become weak and weaker and this is caused by mutation in a gene that is involved in mRNA splicing. So, this protein is involved in mRNA splicing and this protein is having a defect. If it has got a mutation, then it affects the muscle cells; how do they function or survive? As a result, you have an atrophic condition or degeneration of the muscle. Another example is Rett syndrome, one of the common mental retardation syndrome caused by defect in a gene that codes for a transcription factor and what is interesting here is that, this particular gene is known to have multiple first exon, each having different promoter and depending on which one is having a mutation, you would have distinct forms of mental retardation.

We also have another interesting observation in the human genetics that is the presence of long range cis-regulatory elements. What does it mean? What do you mean by long - range, what do you mean by cis regulatory element? So, you have the gene, let us say, this is a region where the gene codes for a protein and immediately upstream of it, you have the classic promoter, on to which transcription factor bind and they transcribe the gene. What is increasingly becoming clear is that there are elements that are present, several mega bases upstream of the gene and these elements are bound by certain proteins and these proteins may cross talk or interact with the promoter which regulates whether the gene should be expressed in a given tissue or not.

Now, such element if this is mutated or some changes happen in this region, may affect the gene from its functioning. For example, one example is SOX9 gene, which is expressed in the

developing testis, as well as in the bone cells. So, it has very different function in these two different tissues and if it expresses in the gonad during development, now or if it doesn't express in the testis during development, then the individual become female. So, it is called as sex reversal. But, if the mutation is such that if affects only the expression of the gene in the developing testis the individual may become sex reverse, meaning become female, although they are genetically male, But if the expression is affected only in the, for example, bones cells or the future bone forming cells they may have some bone effect, but will not have any effect on the testis or male development. So, this happens because of the regulatory sequences which govern the tissue specific expressions of these proteins.

So, that gives you another round of complexity as to how the mutations can affect. So, that pretty much ends the topic as to how changes in the DNA sequence can affect the gene expression or the protein function and we have sort of used the central dogma as an example to discuss how the process of transcription, translation, affect at times the gene function, as a result of the changes in the DNA. That brings the third lecture to its end and we will be looking at gross changes in the chromosome and how they affect the gene function or genome function in our fourth lecture.