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

Module - 01
Lecture - 01
Fundamentals of Central Dogma – Part 1
(DNA, RNA and proteins; mutations)

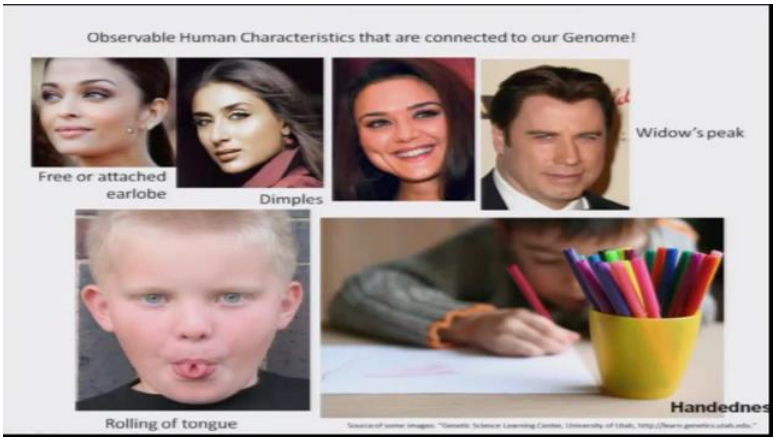
Welcome to human molecular genetics course. So, this week we will be discussing about the fundamentals of central dogma of biology, what you call as central dogma of molecular biology, where we will be looking into the DNA, RNA and protein, how they are connected, how the information is processed and how that brings about the changes, that visible changes that you see and also as to how changes in your genetic makeup may change the way the information is processed. So, that is the theme of this week's discussion. So, we will be looking into some fundamental aspects and I also would give some historic background as to how discoveries have led to the current understanding that we have and I will give some examples of human genetic conditions wherein, defects in a particular process or a particular type of defect can lead to the condition that you see. So, let's look at what really the central dogma of molecular biology.

Anything you see around, in the nature that you see, is a particular shape, form and different variants of that. So, these are normally called as phenotype. For example, in the classic Mendelian form, you have phenotype, that is for example, a plant being tall or being short. So, this is one of the phenotype that we talk about. But, when you discuss about human genetics, there are many changes or forms of expressions that you would see in day to day life.

(Refer Slide Time: 2:06)


For example, I have shown some of the pictures here and there are many interesting features that you may have or may not have noticed. For example, the, the earlobe; how it is attached, whether it is free or attached is one such phenotype, known to be regulated by, for example the genes. These are some of the examples that are shown here. You look around; you see several variants of the phenotype what you call. For example, you have different forms the way the earlobe is attached or people having a dimple chin and so on and that is what is shown here.

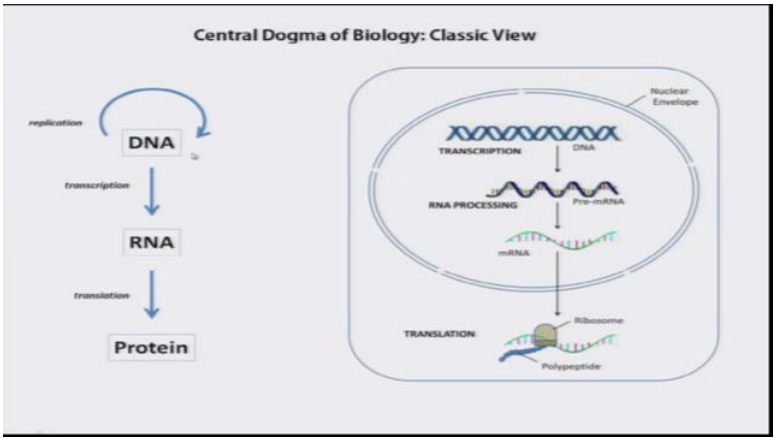
(Refer Slide Time: 2:46)




For example, here you have the attached earlobe, here it is free and you can see the dimple chin and people have very distinct way of hair growth. For example that is shown as widow's peak. Even for example, the ability to roll tongue or not is contributed by the genes. It need not be that all these phenotypes that you see are regulated by one particular gene, but it could be combination of them. But what is understood is, the genetic makeup of your body

contributes to whether or not you are able to roll tongue or whether you have free earlobe or attached earlobe, whether you could have dimple and so on. Even, what is considered is, even the handedness, whether you are at ease using your left hand or right hand even that is believed to be contributed by genes.

(Refer Slide Time: 3:41)

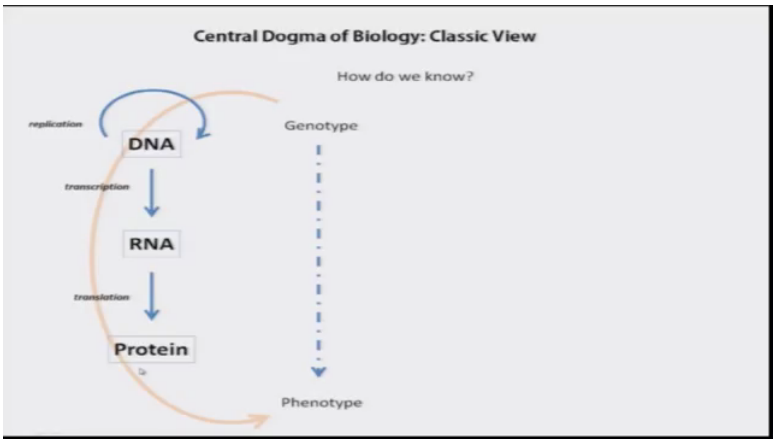



So, that is pretty much what you call as genotype and phenotype correlation. So, if you look into the central dogma precisely that is what is being discussed. So, you have DNA and the DNA has got all the information that we possess and that is coded to RNA and RNA finally gives the information to form what is called as protein and proteins, of course have diverse function as a result you have the phenotypic expression. So, the process by which for example the information that is there in the DNA is carried to RNA is called as transcription wherein, you have set of proteins that help in copying the DNA into an RNA and you have another process which you call as translation which essentially reads the codon that are given in an RNA and convert that into a protein and the protein of course functions.

So, this process is what is shown here in this schematic, where we have our DNA which is there inside our chromosome and that is being copied by the process what you call as transcription and you have this mRNA which is processed to form a matured mRNA which moves out of nucleus into the cytoplasm of the cell where you have set of machineries, that are called as ribosomes that help in reading the RNA to form what is called as polypeptide which on its own can function as a protein or a group of them can join together to form a complex which we now call as protein. So, this is pretty much the schematic which is the classic view of central dogma of molecular biology.

What is also added to that is that the ability of the DNA to copy itself which is very, very essential and this process is known as replication because when the cells divide, they have to carry, each cell has to carry the information and this is, this is taken care by the process that we call as replication. So, how do you know? So we, we talk about the different process; be it replication, be it transcription or translation. How do we know that these processes do take place and how they are regulated?

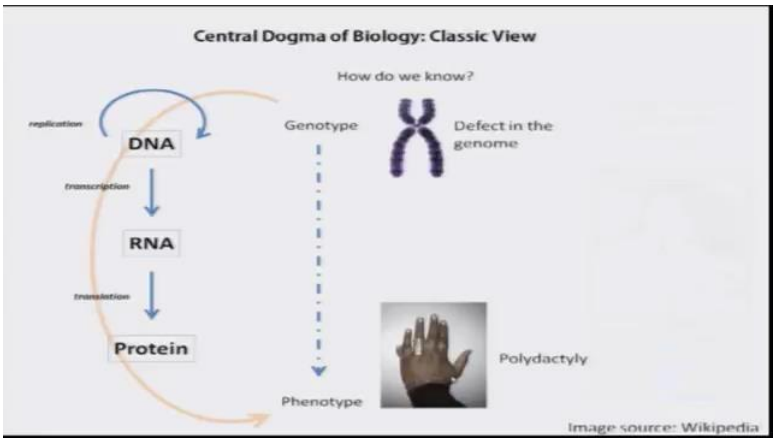
(Refer Slide Time: 6:06)




So, this is something very, very interesting, because classically about 100 and 120 years back people only were able to link these two extreme sets. One you call as genotype which we believe there are some hidden code in your chromosome or DNA. Even then it was not known they called it as factors and many other such names and then, finally what you have is phenotype which is normally looked at. For example, Mendel looked at the size of the plant, the shape of the seed or the color of the flower and so on, while human geneticists looked at for example, the characters that you see. Whether it is for example, your attached earlobe, whether you are able to roll the tongue and, and of course other manifestations like whether you are abnormal, whether you have a disease that runs in the family.

There are many such characters that are looked at and people are able to tell these are, the information for these phenotype or expressions are hidden in your genome what you got. So, that is why, now we know that the genotype is linked to the phenotype probably via this particular pathway, where whatever code that is there in your DNA is, is decoded via RNA and you have a protein which does most of the functions that you see in your body.

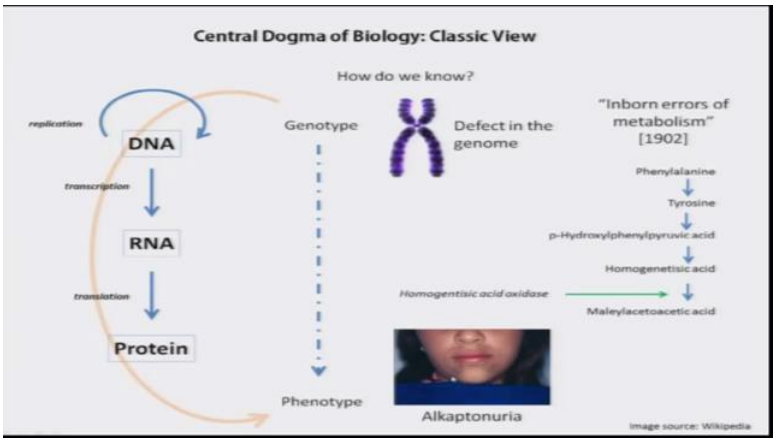
(Refer Slide Time: 7:32)




So, that's how we believe now is the central dogma is very, very essential for all the expressions of all the characters that you see. So, what we know? So, what we talk about genotype is the chromosome. The chromosome carries your genetic material which, about which we will come back a little later, may be in the second or third lecture and this information that is there in the, in the chromosome is transmitted or expressed as a phenotype. What is shown here is a condition that is called as polydactyly, wherein individual could have a sixth finger. So, that is the condition that is shown here and, and this again is a genetic defect, meaning you have had some changes in your DNA which had given a new information which is not desired, which is not normal rather, which is to have an extra finger in your hand which is the sixth finger that you have. So that has come from for example, your DNA to what you have in your hand, the sixth finger. So, likewise people looked into abnormalities, which is, when I say abnormal the characters that are normally not present in the population that is why abnormal. All of us have got five fingers; very few, those who have some changes in their DNA may have the sixth finger. Therefore, that is considered as abnormal.

So, they tried to link and then say whether these are contributed by the genes. One of the pioneering discovery or even what you know as the theory of 'Inborn errors of metabolism' was proposed by a British scientist, physician known as Dr Garrod. He was pioneer, he was a pioneer in understanding, how for example some changes in your genome could contribute to the medical conditions and some of them are involved in the metabolism, for example metabolism of certain amino acids and defect there, how it can contribute to a condition.

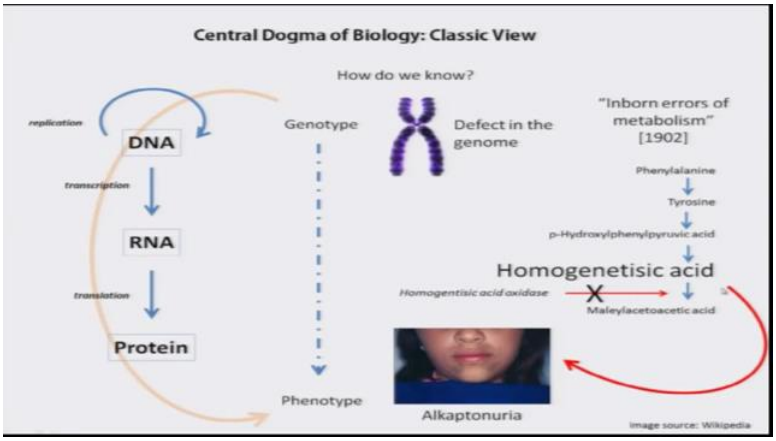
(Refer Slide Time: 9:38)




So, his discoveries have led to the understanding of the fact that way back in 1902, he proposed that you could have defect in your DNA that is what is manifested as a disease which was path breaking and not many have really appreciated that, because they could not believe that. So, what he has looked at is a condition, called Alkaptonuria in which you have abnormal pigmentation like that you see here and these pigmentations is one of the symptoms, but the patient otherwise have several other conditions, which affect their normal physiology. One way this could be diagnosed is to expose the urine to air, a reaction leading to a kind of coloration and that happens because of a a particular metabolite being in a higher concentration and now, what we now know is that in a normal pathway you have the amino acids that are taken from your food, it enters your blood circulation and these amino acids are the building blocks.

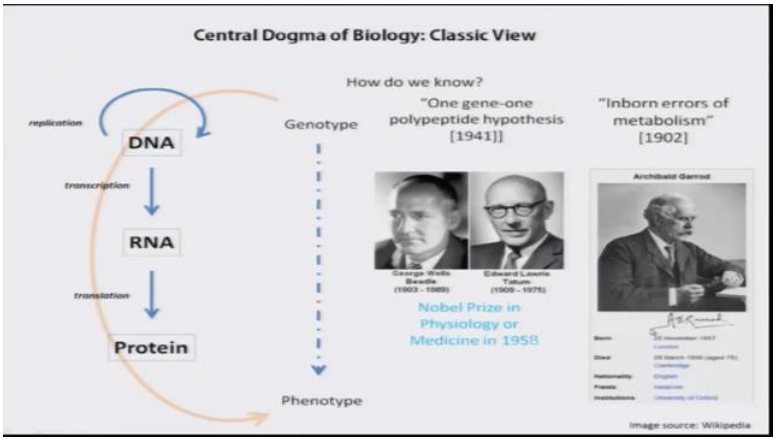
They are being used by various types of cells to form proteins and it is not that all the amino acids are used up. So for example, phenylalanine when you get in a higher concentration, not all of them are being used. Some of the excess would need to be metabolized and, that happens for example in the pathway that is shown here. There are enzymes that convert for example, phenylalanine to tyrosine and so on and you have one particular enzyme that is shown here, Homogentisic acid oxidase, which helps in the conversion of Homogentisic acid into the other compound that is shown below.

(Refer Slide Time: 11:22)



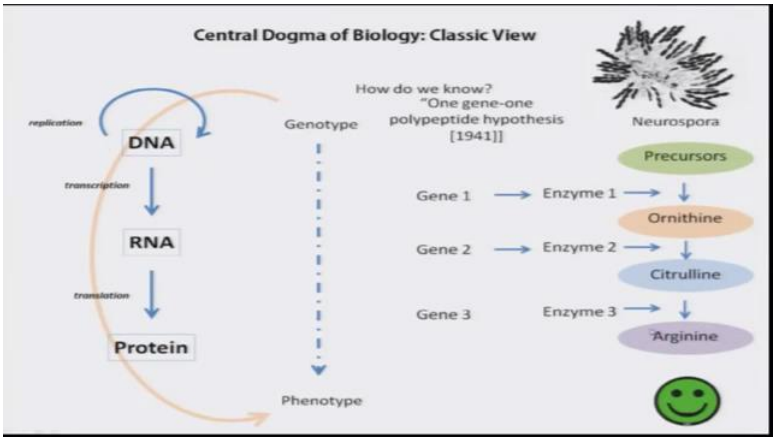

So, what happens in this condition Alkaptonuria is now we know that this particular enzyme or the gene that codes for this particular enzyme is defective. It cannot do the function it is expected and as a result, you have very high concentration of this particular metabolite, at this particular stage, Homogentisic acid which results in abnormal pigmentation and associated phenotype that you see in this condition. So, this is something which is path breaking, I mean that he did not know that there are enzymes involved in different pathways, but what he proposed was that, there is a defect in the genome possibly that alters the way the metabolism happens. As a result, a defective metabolic pathway results in the pigmentation and perhaps that is the first kind of a theory, which suggested that our physiology itself is regulated by genes. So, that we came to know much, much later.

(Refer Slide Time: 12:18)



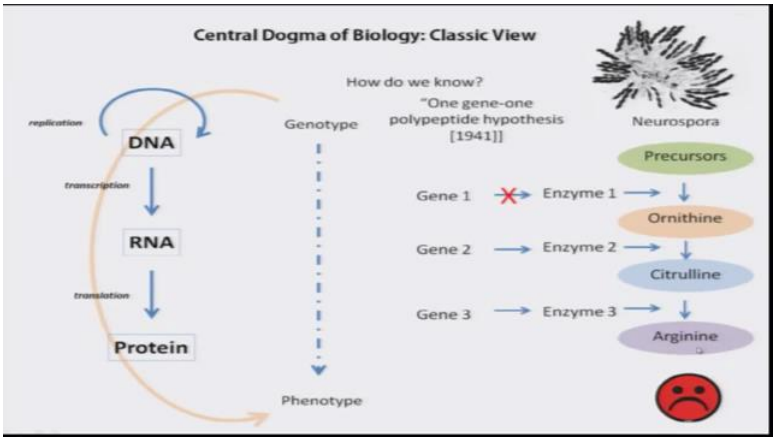

So, in that way he was really a pioneering scientist who contributed much to the current understanding of central dogma. But, his discoveries were not appreciated. It was probably far ahead of the time. It was late until 1940's or early 50's, people really understood that genes are the elements in your body that regulates most of the physiological conditions. This had come from the discoveries by Beadle and Tatum who proposed the hypothesis what is called, one gene one polypeptide hypothesis. They have used very exotic species of fungi and they have shown how defect in a particular gene could change the way the metabolism happens for which they were awarded Nobel Prize.

(Refer Slide Time: 13:21)



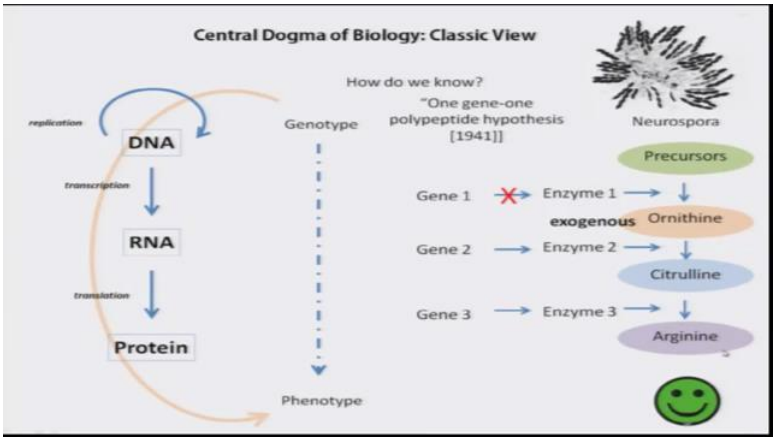

So, this particular discovery have come from a particular species called Neurospora, which is a fungi and what they have essentially looked at is, how for example the synthesis of arginine, or, one of the amino acids happens from the precursors that the plant gets from, for example, its nutrients. So there are again, there are enzymes and each enzyme converts this precursor into the other and so on. This is the pathway.

(Refer Slide Time: 13:48)



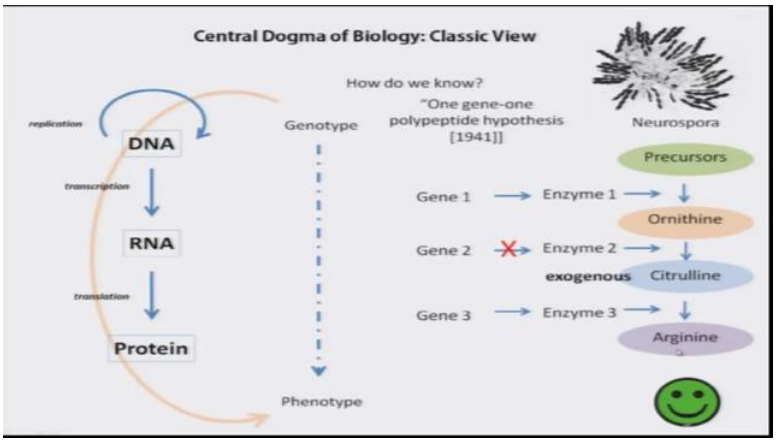

What they have shown is that, if you have a defect in, for example in one of the enzymes, which converts the precursor into the first metabolite ornithine, then again the plant is unable to grow, because it doesn't get the amino acid arginine.

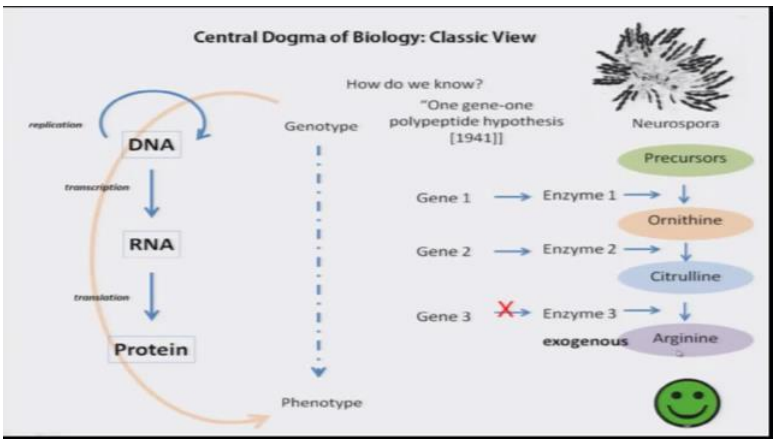
(Refer Slide Time: 14:02)




Likewise, it could have defect elsewhere. But however, they are able to show that if you give this particular compound to the medium exogenously wherein the plant can overcome the deficiency of this enzyme, still then the rest of the enzymes are functioning, therefore it can convert that into arginine.

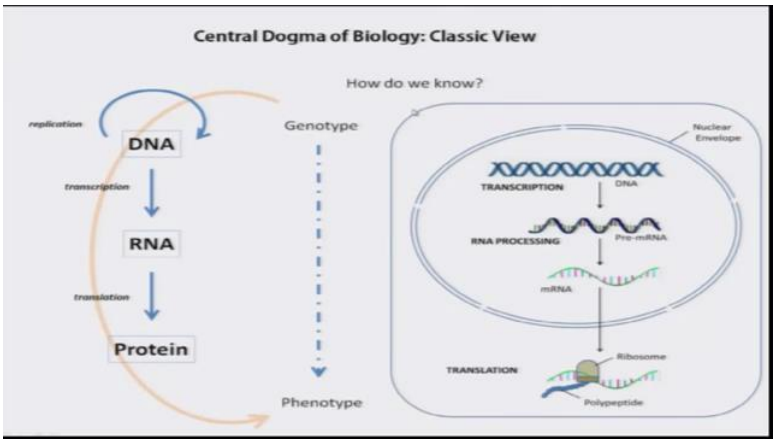
(Refer Slide Time: 14:20)



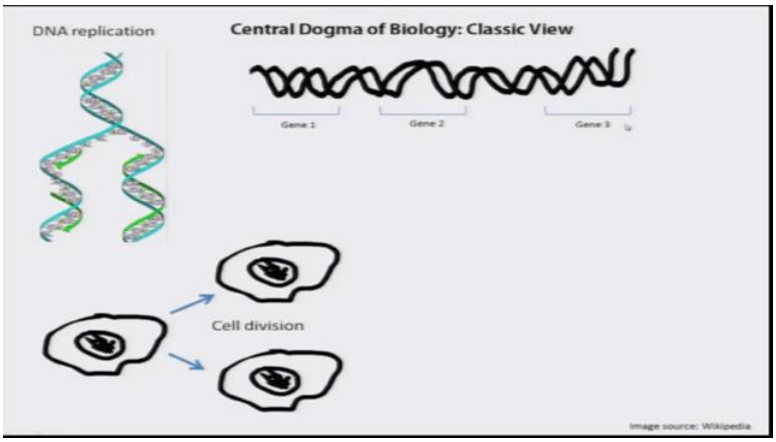
Likewise they have been able to show that there are distinct, genes and mutations or changes that affect how the enzymes metabolizes these compound and then, therefore they are able to get the amino acid which helps the fungi to grow. So this hypothesis which you now call as one gene-one polypeptide hypothesis was proven by these experiments and which led to the discovery that we talk about that all these metabolic pathways and other events that take place in our cell and body are regulated by genes.

(Refer Slide Time: 14:55)



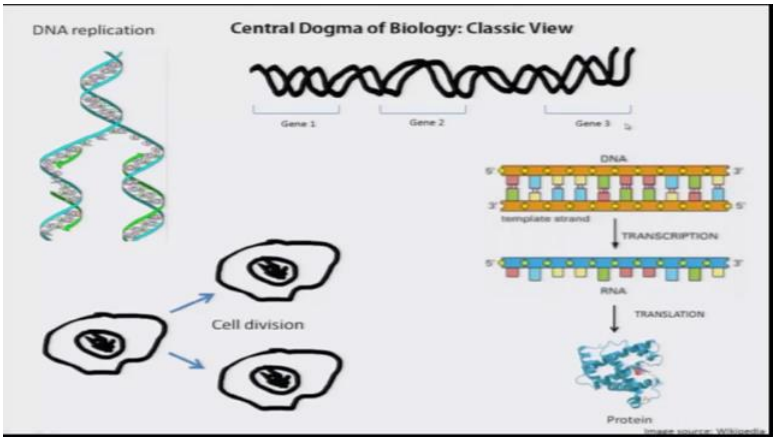

So, that is the foundation for our understanding of central dogma of biology; DNA to RNA to protein.

(Refer Slide Time: 15:13)



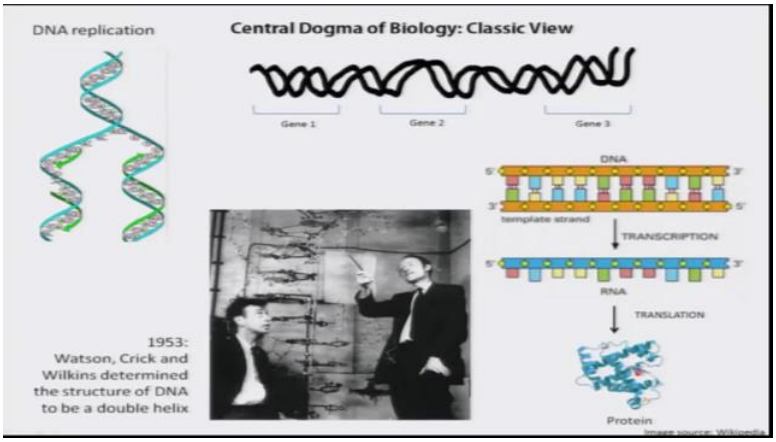
So, in this process what is important is the replication, because when the cells divide, the cell has to make sure that it copies its genome without any error, because this cell, gives rise to two daughter cells and these two cells are identical to the mother cell in a normal mitotic process and this happens by the process of replication wherein, the two strands of the DNA unwinds and then the machinery copies the two strands. So, when we talk about the DNA, we say that the DNA is a component of the chromosome; DNA plus protein form the chromosomes and that DNA runs from one end of the chromosome to the other. Also, the DNA has got several genes. So, it ranges from thousands to hundreds, depending on which chromosomes we are looking at, which we will be discussing little later. But, what we know is that this linear DNA molecule has got several distinct regions, each one of them could have the code or signal for certain protein and that is say for example, protein1 could be coded by this segment of the DNA, therefore you call that segment as gene 1, gene 2 and gene 3 and so on.

(Refer Slide Time: 16:30)



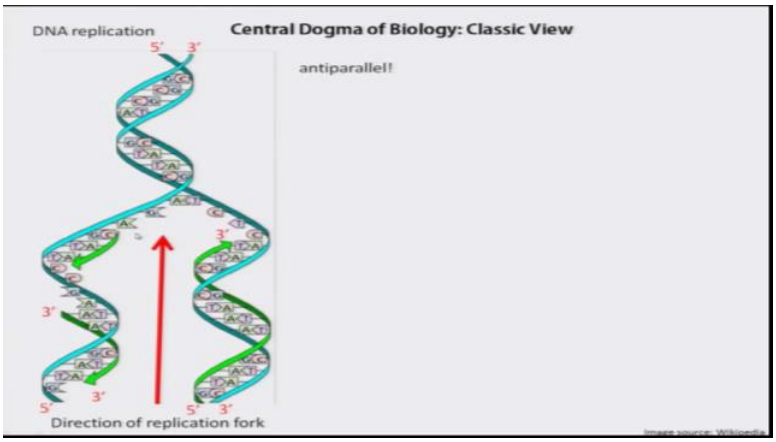

So, this process happens via transcription and translation, so meaning that you have certain machinery that binds to this region of the DNA and then copies the DNA and that is what you have as RNA and the RNA is copied into or translated, decoded into a protein, final product that you see.

(Refer Slide Time: 16:50)




So, the very concept, that the DNA is double stranded and they run antiparallel to each other was proposed by Watson and Crick in 1953, for which they were awarded Nobel Prize and now we know that it is no longer a model; that is what we know is the true status of the DNA that it is antiparallel; it is complementary to each other it is universal, and that it has got the four basis, is considered to be the rule now.

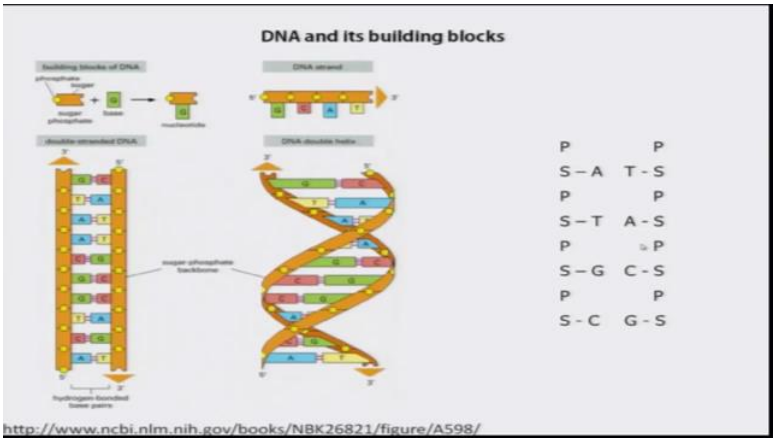
(Refer Slide Time: 17:24)




So, when we talk about replication that is cell copying its DNA, therefore the DNA can be given to the two daughter cells, this process we call as DNA replication and you can see that, in this cartoon what we are showing is that this is the direction of the replication fork. Thus in this process the two strands are separated and a new strand is being made and why the DNA

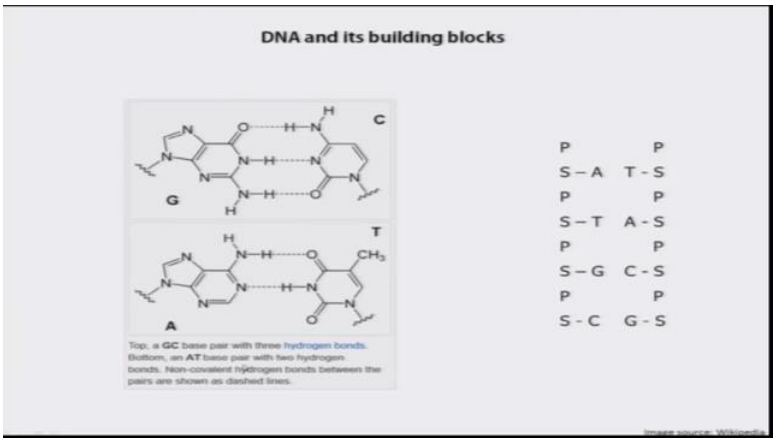
should be antiparallel, because, it is also a challenge for the DNA replication machinery, since on one side it is able to make a continuous synthesis of the new daughter strand, but on the other strand you do have a challenge that is to make bits and pieces of them, because of the direction of the formation of the polymer; always it is 5' to 3'. But still we don't know, because there could be evolutionary reasons, so we still don't know as to why the nature has selected such anti-parallelism. What it explains however is that, by having the two strands, the DNA is able to copy.

(Refer Slide Time: 18:36)



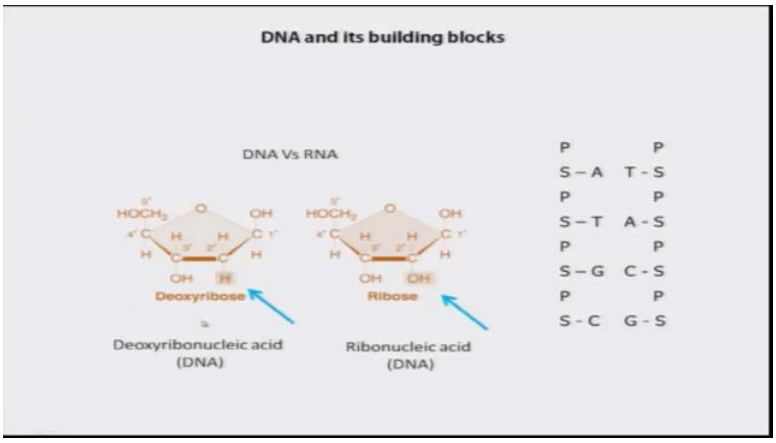

There is what is called as complimentarity, wherein one base pairs with the other. That's the way either of the two strands can serve as the mother strand and make a new daughter strand. So, that is what is shown here. So, the cartoon gives you a kind of understanding that probably most of you know is that the DNA is made up of these three components: a sugar, a phosphate and the base that together forms the nucleotide and the polymer of that is called as a DNA strand and two such DNA strands, hydrogen bond and then forms the helix, something like what is shown in this side - that you have a base, A pairs with T and G pairs with C. So, this is universal. Now, what you have at the outside is the phosphate backbone, and then you have the bases and this is pretty much similar in other context as well.

(Refer Slide Time: 19:33)



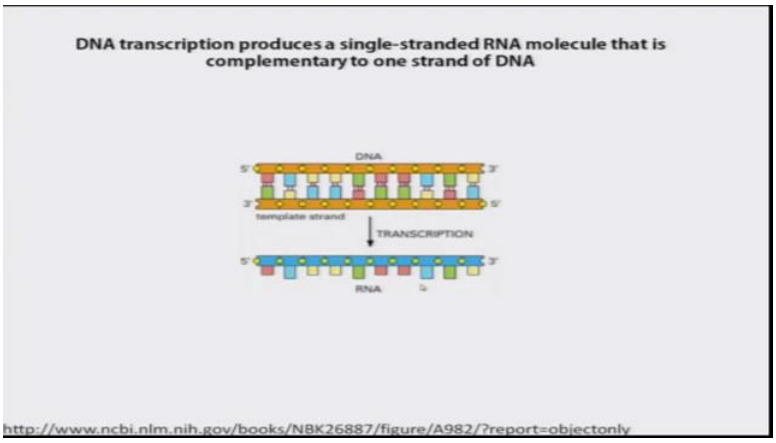

But let us look into the bases, wherein, C pairs with G and A pairs with T and each one of the letter refers to the base cytosine (C), adenine (A), thymine (T), guanine (G). So, that most of you are aware of.

(Refer Slide Time: 19:52)



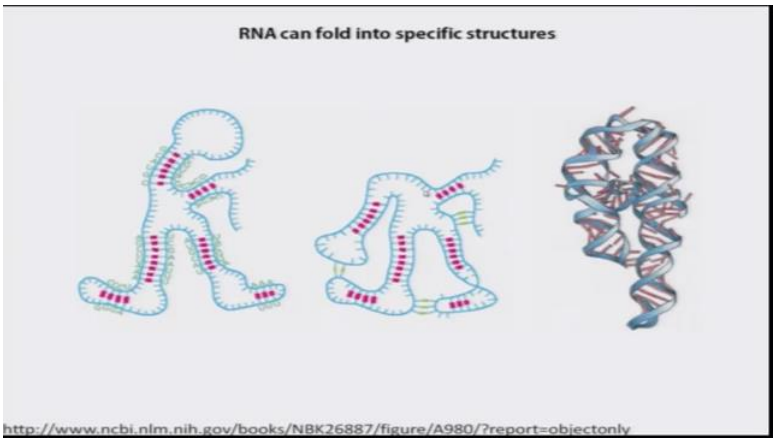

What is also interesting is that the bases are pretty much identical between RNA and DNA, except that, in this particular position the ribose has an additional, oxygen in RNA, while it is absent in DNA, therefore DNA is called as a deoxy and that distinguishes the DNA and RNA. The difference is that, in the DNA in this particular position you do not have the oxygen, whereas in the RNA you have an oxygen moiety here. That distinguishes these two elements.

(Refer Slide Time: 20:36)



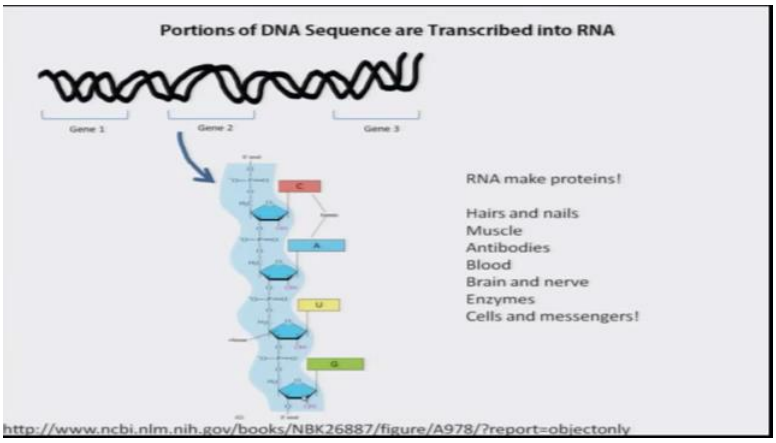

Otherwise they are identical and they can base pair and form like what is shown here. So, you have a transcript which pretty much, copies the DNA and that is how the signal is, is copied or the code is copied.

(Refer Slide Time: 20:46)



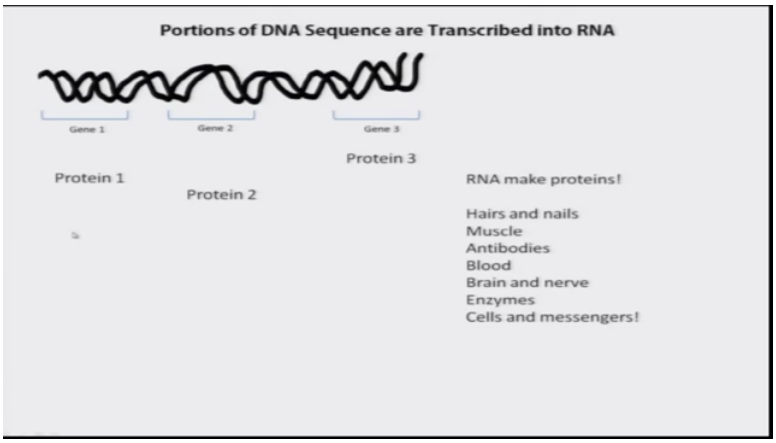

The RNA, these are single stranded forms as most of you know and they do form subtle structures and the structures gives them the structural and the functional property. For example, tRNA or ribosomal RNA and many other RNA's, they function with the structures that they get because of the sequence that they have.

(Refer Slide Time: 21:04)



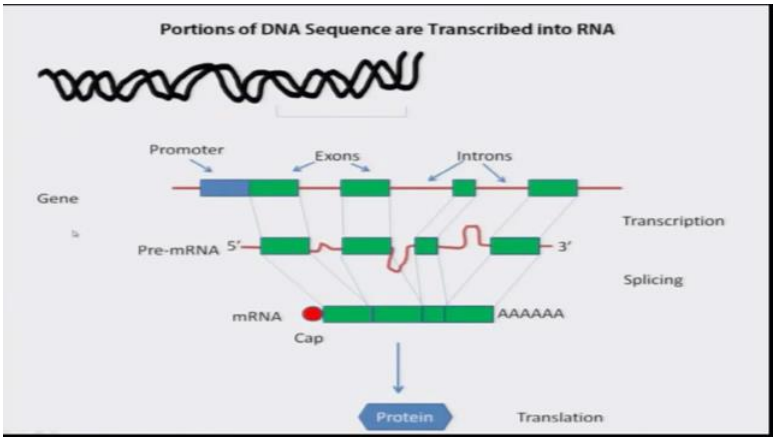

Now as I explained, a portion of the DNA sequence is transcribed into RNA. So, that you call as different segments, which are, gene 1, 2, 3 and so on and one of the segments is copied into an RNA and that would have the complimentary sequence of this strand that serves as the template. So likewise you have C, A, U, G and so on, you have the sequence that are copied to the RNA and this RNA helps in the formation of proteins, because they give the signal as to what kind of protein should be made and, meaning in what combination these amino acids should be added. So, that provides the signal for making the protein and you have proteins everywhere; whether it is your hair, nail, muscle, antibodies, blood, plasma or even the different types of the cells that you have, brain, nerve, enzymes and you name it, any of them would be made up of, proteins. So, that signal comes from your DNA via the RNA.

(Refer Slide Time: 22:17)




Now likewise, some gene could code for an enzyme, the other gene could code for a transcription factor, some other could code for a protein, hormone and so on.

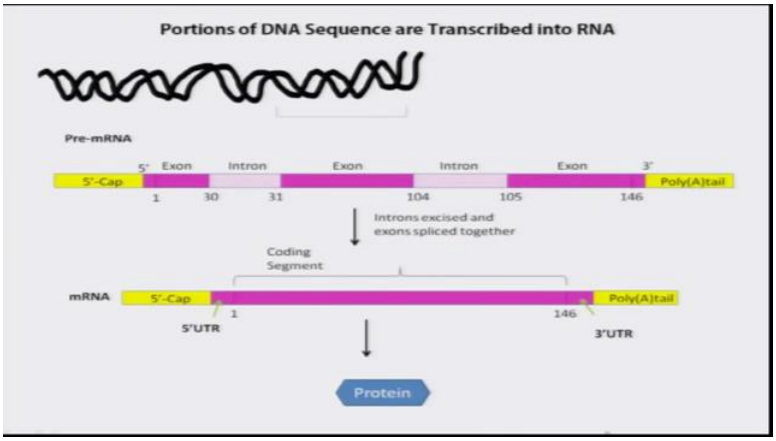
(Refer Slide Time: 22:28)




So how this is, the coding or, the information from the DNA to protein is passed on? So, the way it is done is that every segment, for example if we look into this particular segment of the DNA and that you call as the gene here and this has got an element called as promoter. So, this element is nothing but a signal sequence for certain transcription factors to come and bind and as a result, the region from here to here, for example in the DNA, is copied into an RNA. So, what it does? It basically copies the sequence such that, one of the strands is faithfully copied it into an RNA strand. Now this, during this process when the RNA is made which you call the pre-mRNA, this RNA undergoes certain changes and this process is called as splicing. For example, there could be regions which are present, shown here in the red color line are removed to form what is called as matured messenger RNA.

So, the regions that are retained in the mRNA are called the exons and the regions that are not retained, or removed during this splicing process are called as introns. So, there are many genes that has got multiple exons; 20, 30 and so on. There are few genes that do not have introns meaning they have only 1 exon. So, there are exceptions. But as and when a particular gene has multiple exons, these regions which you call as introns are removed by this process called as splicing. Once that happens, the mRNA gets modified. There is a cap, modification that takes place at the 5' end of the RNA and then there are poly A sequence added at the 3' end of the RNA and this RNA gets to the cytoplasm where it, the signals or the information is processed and then you form what is known as protein via the process translation. So, this happens using a very dedicated set of machinery which involves ribosomes, tRNA and many other factors.

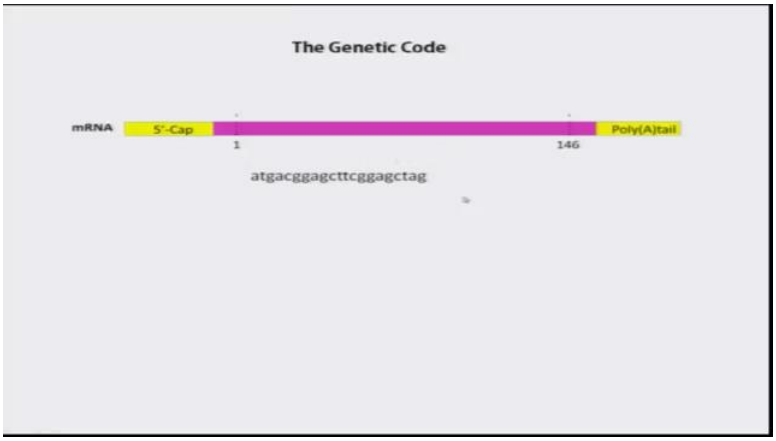
(Refer Slide Time: 24:55)




What is interesting is that, this process you have, how this intronic element which is normally removed in the RNA, how that particular removal or what you call as splicing takes place? So, when you remove this region and when you join these two regions, that is what is shown here for example, you have the matured RNA, which has got the 5' cap side and then you have the poly A tail and here, the number 1 to 146 for example, denotes the number of amino acids this particular RNA could code for and what you have is that upstream from this is the code for the first amino acid, but still you have certain sequences present upstream of which, which is known as untranslated region that is what is abbreviated as UTR.

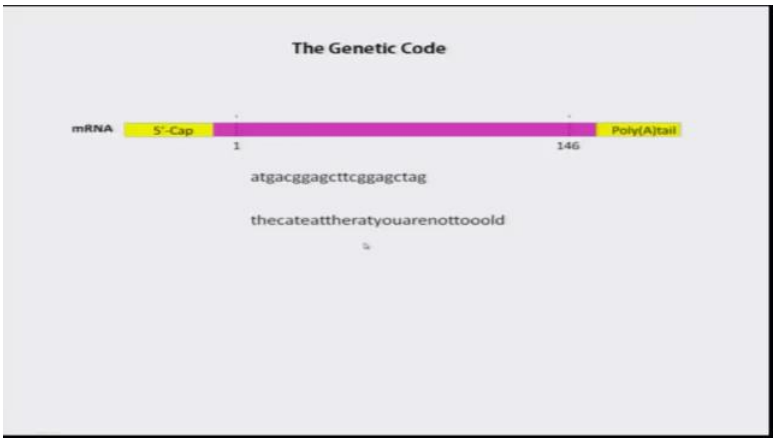
Likewise, this is the signal for ending the peptide and you have sequence beyond that until the poly A tail, which you call as 3' untranslated region. So, the entire region including the region that represent UTR are known as the exons. So here, what you see is that even this UTR is known as exons, because these are the regions that are present in the gene and they are retained in the mRNA. So any region that is not retained in the mRNA is known as, called as Introns. This is one of the confusion that we often find in many of the students. They believe that only the coding region or the region that give the information for the amino acid is called as exon, which is wrong.

(Refer Slide Time: 26:39)



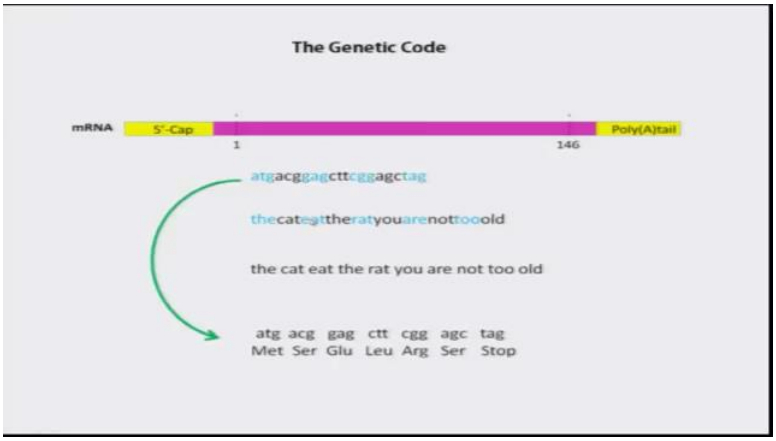

So, what is a genetic code? Now, let us look into how the sequence in the mRNA is understood or decoded. For example, what we know is that, the sequence that is present in DNA has got the combination of the 4 bases and RNA of course you would have almost same, except that you would have, U in place of T. But for most of the sequence that you derive from DNA therefore, I am trying to show the same 4 bases and this is a kind of sequence that you see. The combination of the 4 bases A, T, G, C. Now, how would you make any sense out of it as to what really the message is? That is really difficult.

(Refer Slide Time: 27:23)



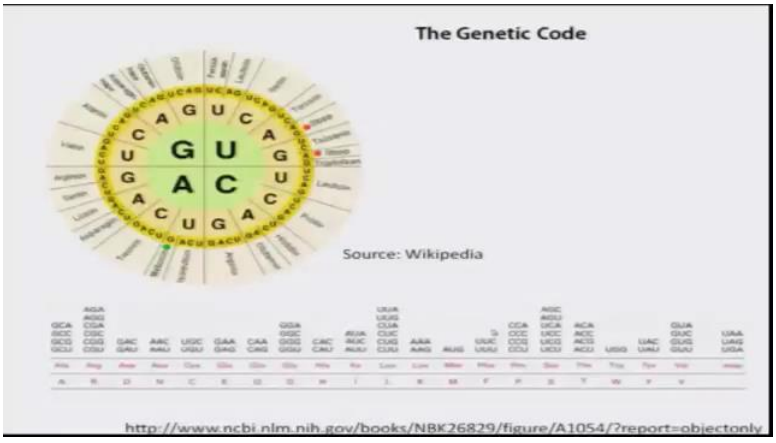

It is not if you know the rule. For example I have scribbled some, letters here, not necessarily the one that represents the DNA, but the alphabets that you use for, writing a sentence, letter, e-mail and so on. The very look of it you will not able to understand what it is.

(Refer Slide Time: 27:42)




But if you can carefully look at it and for example I colour coded them into a group of three letters, then it makes sense. Now I can read. The cat eat the rat and then you have another sentence, you are not too old. So, this gives some information. Now I can understand this. So, if you are able to group the three letters together, you are able to read out what really the meaning of these letters. Exactly the similar way, it happens even for the sequence that is present in the RNA. For example, if I group them into three letters, for example atg, followed by acg, gag and so on, now I know, for example the atg codes for the amino acid methionine. Likewise, each of the group having three letters codes for a different amino acid, the last one being a signal that the peptide should stop here. So, certainly now you can make the sense out of the sequence which otherwise was challenging to begin with.

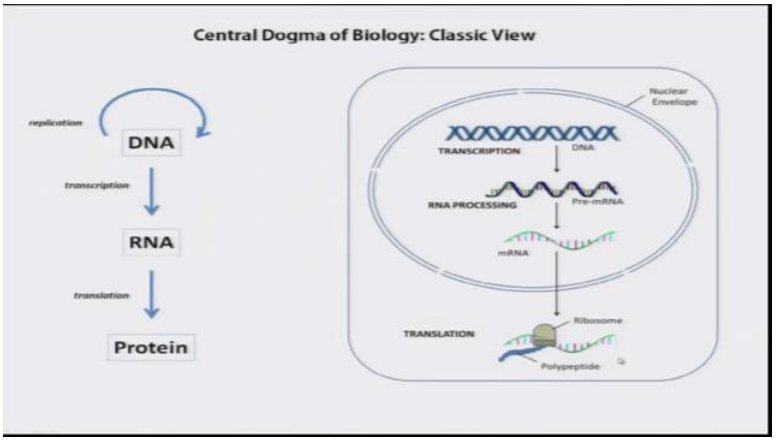
(Refer Slide Time: 28:44)




So, this is the code that the 4 bases that you have - the A, G, C and U that you find in the RNA, can have as many combinations that are shown here and each one of them have certain specific code. For example, four different codes of triplets give the same message that it has

to make alanine. For example, the AUG triplet has got, the message that is methionine, it has to code for this particular amino acid. I have three different codes, UAA, UAG, UGA, all three gives the signal that the peptide synthesis or protein should terminate here. So this explains as to how your cellular machinery can read the RNA or sequence that are embedded in the RNA and make sense out of it. So, this is exactly shown in two different ways. One is to show all the combinations. For example A combines with G and that can be combined with any one of the four and then you can see how it gives the different codes which you call as codons. These are the triplets, or else you can see that these are the amino acids that we use for making the protein in our body and you can see how many codons are required for each of these amino acids. Some use more than 4, some use two and a few of them use for example only 1 codon. So, that gives information as to how in our cellular machinery, can read the message and make protein as we desire.

(Refer Slide Time: 30:34)




So, that pretty much explains as to how a DNA which copies itself to give a copy of it to every cell that is being used to make an RNA and how RNA, after all the processing that you spoke about like splicing, capping and polyadenylation, how that helps in making the peptide which you call as a protein and which gives the phenotype. So, this is a basic introduction about the central dogma and with that we will end the first lecture. In the second lecture we will be talking about more information as to how changes there can affect the way the cell functions.