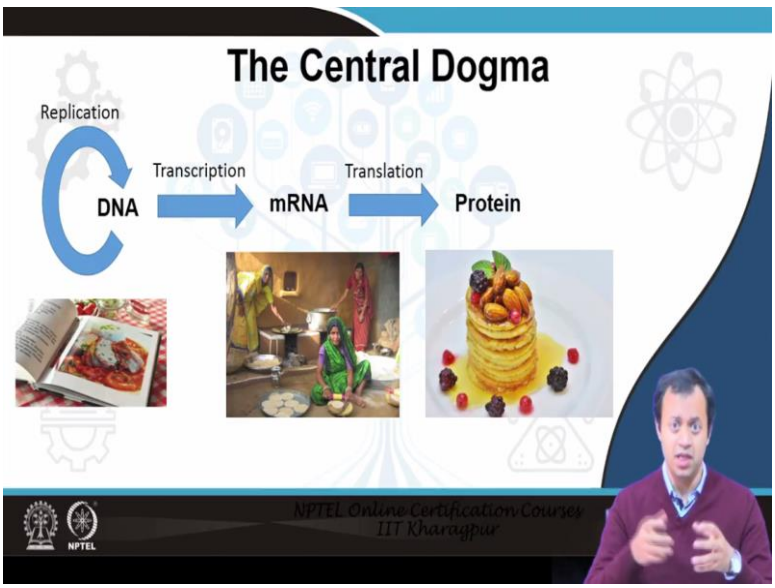
Biophotonics
Professor Basudev Lahiri
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur
Lecture 13
Genetic Code

Welcome in the past few classes we discussed about biology, the very basics of biology and we also discussed the central dogma. The central dogma is one set of rule, one set of protocol that forms the heart of molecular biology. I even asked you to look into central dogma a bit detail. Central dogma very, very rudimentary, very, very simplistically states that DNA creates RNA, RNA creates protein.

But obviously there is huge amount to unpack in here people spend their lifetime simply to understand to elaborate just one part of the central dogma. As I said my metaphor was central dogma is a combination of Schrodinger wave equation, Einstein's theory of relativity and the Maxwell's Electromagnetic laws all combined into one and it forms the bases of molecular biology.

We will not try to go much deep into the nitty gritties of central dogma, but since we are a technology-oriented course we will try to look at the manifestation, we will try to look at what is the output, what is the outcome, what is the culmination, what is the result of the things that we achieve through central dogma. So, welcome to the class on biophotonics. We will continue with our course on basic of biology and today we are going to understand the genetic code. It is going to be a bumpy ride so stay put.

(Refer Slide Time: 01:54)

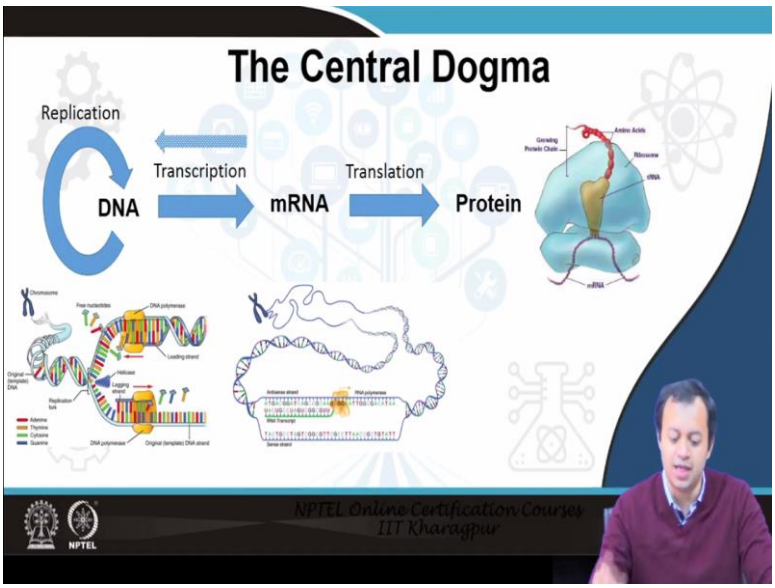


So, last time I discussed the central dogma DNA is basically information, it is the recipe book, it is the idea in recipe book or in the circuit diagrams you have the ideas of different circuits everything in between. RNA is the actual process of cooking. In the process of cooking you require several different utensils, you require several different procedures, you require the heat.

You require maybe certain spatula, you require different types of tools and final product, the final product which is the dish which you all consume is the protein. This is the overall very basic, very rudimentary metaphor, but remember there is unlimited amount of information that can be unpacked in each step. Scientist spend their lifetime understanding how DNA get converted to RNA and then RNA has several different formats.

I am talking mostly about messenger RNA then messenger RNA along with tRNA forms protein in ribosomes all these are quite a detail endeavor and I have asked you to look into the animations which is one in my opinion one of the best ways to understand what exactly is going on in between. I am going to skip the entire detail how DNA gets converted into RNA and how RNA gets converted into proteins for all the bad reasons.

(Refer Slide Time: 03:34)



But I will just give you a glimpse of what exactly is it. So, replication simply means that the DNA strand the double helical DNA strand opens up just like a zipper it opens up and both forms this side and this side they individually form their own DNA. So, we call the DNA replication in which DNA converts or DNA duplicates into another sets of DNA. One DNA gets converted into another pair or another sets of DNA that is called replication.

You can consider it as duplication itself. So, replication simply means DNA gets converted to DNA. One DNA produces another pair of DNA. Each of the double helical strand breaks up this makes one individual DNA, this makes one individual DNA containing parent DNA one strand of this parent and the remaining this parent and the remaining part is new. So, we say DNA replicate semi conservatively meaning one part of the DNA remain its own parental DNA.

Another part is a new DNA which is the daughter DNA. Obviously, there is lot to unpack here how one strand is faster, one strand is slower. I let you do it at your own pace. This is not part of this syllabus so I am going to skip through it plus it will take me one hour simply to teach you what replication is. When DNA gets converted into RNA we call the process as transcription.

You understand by this time that the difference between DNA and RNA is simply that the thymine base pair is removed as Uracil base pair. So, here the T is removed and U is put again the two different strands they are opening up and each strand is coming together with groups of enzymes. There are RNA polymerize which converts DNA into RNA then there is also DNA polymerize is which is another enzyme which helps DNA to duplicate.

So, these enzymes are nothing, but proteins in itself. So, a specific type of enzyme RNA polymerize helps in forming RNA from DNA. Finally, we get protein this mRNA strand goes to the ribosome. Remember ribosome we discussed along with tRNA it combines together and forms amino acids. Amino acids combination of several different amino acids gives me a long polymer chain, the polymer is basically a protein.

The polymer containing large number of amino acids is protein. This is the overall system huge amount of things needs to be unpacked just do me a favor look at some of the YouTube videos, some of the cartons and you will see how fascinatingly this is done. We will not discuss the entire detail of transcription, replication, translation, etcetera. We are going to understand so what DNA has converted to RNA.

RNA has converted to protein so what and whatever has been the output can be utilized it for our any own goal. Now, here it needs to be understood that there is a process called reverse transcription where RNA can be reconverted back into DNA. However, that is almost exclusively laboratory based in natural processes. DNA gets converted to RNA, RNA gets converted into protein granted.

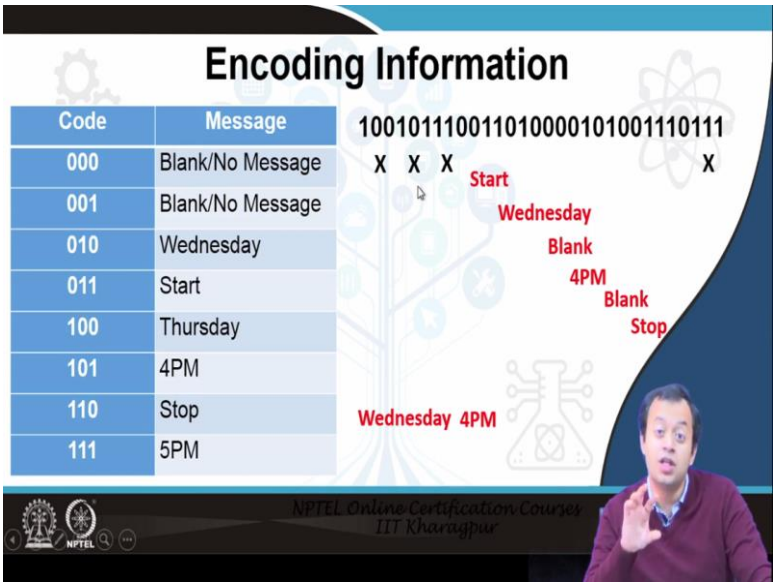
Certain circumstances RNA can reconvert back to DNA, but once RNA has been converted into protein once you have made your dish, once you have cooked your food, once you have made your cake or biscuit or roti or chapatti you cannot break it down into a recipe book think about it. The recipe book gets translated into cooking whatever the processes of cooking going on.

You ask someone to write it down it reconverts back to recipe book, but once the food is made, once the food is in your dish that cannot get back into the recipe dish, wrong metaphor or unclear metaphor, but still a metaphor nonetheless. We are going to understand what is gene in here. Here, there is no mention of anything called gene or genetic code or genome or any such thing.

So, where is gene situated in anywhere of this central dogma and what exactly is that you all have heard of gene we keep on talking that is genetic I have got my father's eyes, I have got my grandmother's hair color etcetera and that is gene that gene is going on what exactly is gene and how are the genes related with DNA and mRNA and protein and all of that. Yes, I can hear the medical student chuckle in my ear.

But remember this is scattered specifically for those people who have not studied biology for a pretty long time. So, the idea is DNA encodes information that information gets converted, gets synthesized into protein, but what does that actually mean, what exactly is the meaning of the term encoding information?

(Refer Slide Time: 09:19)



So, let us go into this part I will tell you a story to understand what exactly I mean by encoding information that day I was reading a book a spy thriller by Frederick Forsyth called the fourth protocol. They even made a movie out of it starring Pierce Brosnan the guy who played James Bond. So, the plot of the story it is a soviet spy, KGB spy has been sent into England where he is hold up where he is staying there as an Englishman disguised and he is receiving information.

Upon receiving the information, he goes and he meets a person that person gives him certain ingredients, certain tools, certain equipments, small equipments one bit at a time. He then comes back to his home in England and with this equipment builds a nuclear bomb so that is the premise of the story of Frederick Forsyth fourth protocol. Just watch this there is even a movie associated with it I think you will love it if you are interested into spy thriller etcetera.

So, I am going to use that metaphor into how information is coded in DNA if you are curious follow me. So, the information was given the spy was given this information that he will receive streams of bits containing 0s and 1s and these 0s and 1s they will contain messages, but there are certain rules what are the rules pay attention.

The rules are the message will be always in triplets three bits at a time, three bits will have messages not all of the bits will have message, some of them are blank. So, this is just for security purposes then there will be a bit coming which will have a start bit and then there will be bit which will be stop bit. For additional security, I will give you a long stream of bits, long stream of messages, but the message in that entire bit will only be contained into one starting with start and ending at stop.

So, anything other than start and stop is invalid understand it. So, I will send you a long bit of information you will decide for them three at a time triplet by triplet and you will try to find out what are the information in here not all of the messages will have information and the information will be valid only when at the beginning there is a start triplet and at the end there is a stop triplet.

If you receive low start triplet then this is a random message this is just send to confuse the enemy. At the same time if you have start and stop and there is mostly blank messages that is also an invalid message done to confuse the enemy rules are clear. So one day the person got this message, this bit. Now the spy has to decipher one triplet at a time what is the message if at all present in this stream.

So, the very first thing he needs to do is to look for the start bit. So, does 100 consists of the start bit. Well, no, so it is an invalid bit irrespective of whatever be the message encoded. If it is not start bit it will not be valid. Similarly, 101 is it start bit no invalid. Similarly, 110 is it start bit no invalid. What is next? 011 okay 011 start so my message is starting. So, my message is starting.

What next? 011 has come then I have 010 starts with Wednesday next message nothing blank. Next message 4 PM next message blank, next message stop and after stop it does not matter I have made invalid. Only the message which is between start and stop will be valid there might be blank messages in between for security purposes and there will be streams of data which will be outside start and stop which irrespective of the information will be invalid.

Meaning even if you have received this part somehow this has been stolen or someone has been able to do it, but if we do not know the rule that the message is only valid between start and stop bit the message will not be decoded. So, the overall message from this stream is Wednesday 4 PM meaning at Wednesday 4 PM the spy is supposed to go to meet the person the person will give him the package the package will contain a tool.

He will then use the tool to assemble a nuclear bomb nightmare scenario I know. So, what exactly has this anything to do with DNA coding, what it has to do anything with DNA coding valid question. Imagine what is gene, what is genome, what is DNA anything. If this entire stream of data 10010111 this entire stream is DNA then this part the one which is between start and stop.

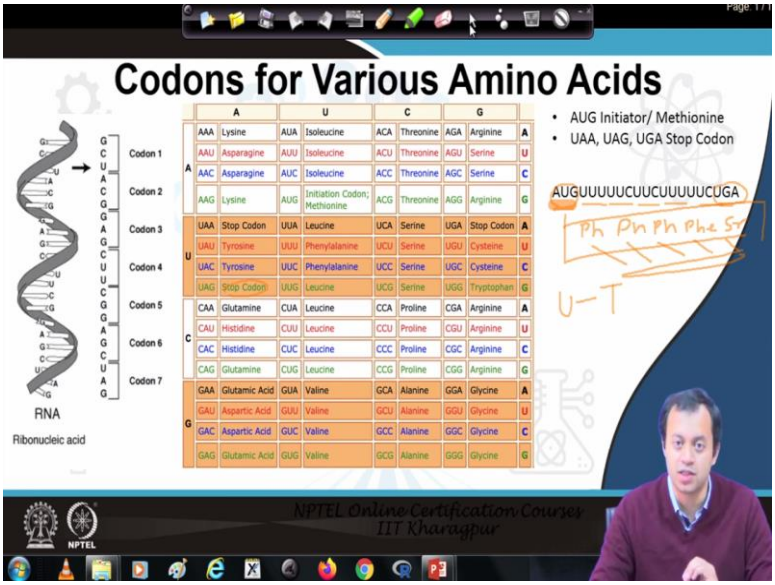
The one which got encoded, the one which got beg your pardon decoded, the one which actually give you a meaningful information which actually gave you something to do is the gene. The entire thing the streams is DNA, gene is a part of the DNA which can be synthesized into protein. Now make no mistake it can be synthesize with protein, it does not mean that it will always be or there will be also parts which are blank which will not be, but still something between start and stop will get converted into protein and that is gene.

A gene is a functional unit of hereditary. It is a part of DNA that which can get synthesized. Now other parts of DNA can also get synthesized in longer period of time somewhere else, but for a given period, for a specific stretch of DNA it has to start with a start message, it has to end with an end message. The start message we call it promoter, the end part we call it or the stop part we call it terminator.

And everything in between can be considered as exons and introns. Exons are the messages Wednesday 4 PM, introns are the blank messages that does not go. Trust me this is the best way for me to teach electronics engineers or physics people how gene encodes information. I know medical people have learned it completely different way obviously not with a truth table with 3 bits, but I am utilizing 3 bits because every electronics engineer, electrical engineer physics person knows this code.

I am simply translating it. Now again make no mistake this is not exactly how it is happening I am simply giving an analogy, no analogy is perfect, there is so much to pack thing, but for all intended purpose if you are looking for the first time into genetic code consider this. The valid information within the stream is gene. The valid information within the DNA is gene. Gene is a part of DNA that actually gets synthesized into proteins and you see the remarkable similarity of this with how things actually happen.

(Refer Slide Time: 17:43)



So, just like the information was coded in bits of 1s and 0s it was a cosmologist, a Russian American cosmologist called George Gamow who decided that if there are four base pairs and 20 amino acids you require only three base pairs at a time to encode information which is protein. Just like previously 100 was giving a message, 101 was giving another message, 100 was giving another message.

Similarly, he decided or he figured out along with Crick as such that three base pairs at a time triplet also known as codons can be sufficient to create all different amino acids. There are 20 amino acids actually 22, 20 amino acids which are directly converted from RNA.

Amino acids combination of several different amino acids forms proteins, proteins perform all the life processes in our body reproduction, respiration, immunity, our cognitive abilities, our ability to move, our ability to talk, our ability to digest food.

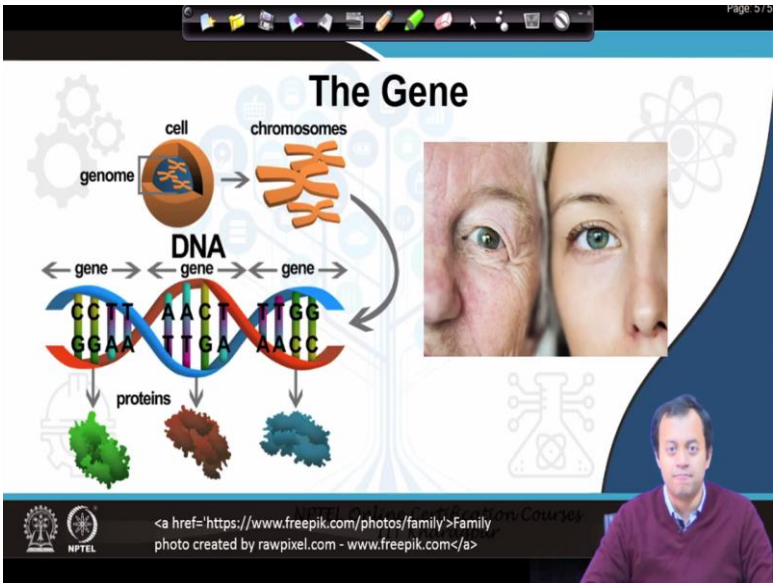
Our ability to grow all of these are determined in directly or indirectly by proteins and what proteins will be produced by our body are directed by the gene. The gene contains these triplets this is very, very simplification I know genes contains triplets, triplets are the base pairs, three base pairs at a time and they are there to produce group of amino acids. So, just like in previous stream remember the previous slide you have a start codon an initiator.

AUG is the start codon. So, if I now decode the message in here AUG is the start codon then you have UUU triplets three at a time just like previous case. UUU is Phenylalanine. So you have Phenylalanine then you have UUC which is Phenylalanine as well then you have UUC which is Phenylalanine then you have UUU which is Phenylalanine then you have UCU which is serine and then UGA is your stop codon.

UGA is your stop codon. So, this gene containing this kind of base pair is going to give me this protein that is the overall information. Do me a favor when you are at home try to see all of these base pairs if you are looking into DNA just replace U with T otherwise this chart remains exactly as it is. Look at any of these ATCG, ATCG, AAAA stream and then code what kind of protein you can create out it.

What kind of protein will Phenylalanine and serine will create. These 20 groups of amino acids give rise to plethora of proteins that is overall the beauty of it. So, again gene is a part of DNA which contains information which could be synthesized into proteins and we have triplets of base pair called codons and these codons each of them one codon synthesizes amino acids. Just three codons or just one codon that is three base pairs can create one single protein.

(Refer Slide Time: 22:09)



So, let us move on. So now I think we have some understanding of what genes are. So, you have chromosomes present in the nucleus, we have 23 pairs of chromosomes we get each pair one sets of pair from our father, one sets of pair from our mother. These chromosomes contains DNA, some part of the DNA which can be actionably converted into proteins are basically the gene.

So, depending on all of that you have your same color of your grandmother's eyes because the protein that is coloring the pigment of her eyes get passed into you via the gene and that gene present in your body is coding or is producing or is synthesizing the same amount of proteins, the same type of proteins, the same nature of proteins which give rise to the exact same color with exact same definition.

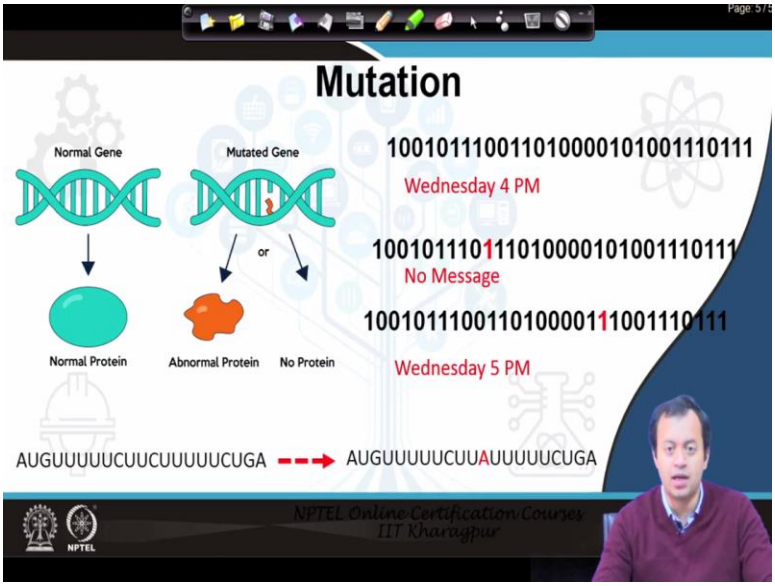
Now, remember this thing and this is absolutely, absolutely fascinating. The last century we were able to understand the fundamental irreducible unit of matter and that is atom some of you will say that atoms is not irreducible, atoms can be broken down into electrons and protons granted, but atom is that one thing that gives rise to matter. If you take a matter you start breaking it down the final product which you will get is atom.

From atom the process could be reversed you join one more atom you have molecules, you have set of molecules and then you have matter, but when you break down the atom into electrons and protons bunch of electrons and bunch of protons by themselves do not create an atom and do not create a matter. So, for all intended purpose the irreducible unit of matter is the atom. Similarly, the irreducible unit of information is the bit either 1 or 0.

You know in a long stream of data made of 1 and 0 if you change 1 it changes, matter is made up of large number of atom if you change the atom the matter itself changes. Similarly, the human being the information present in a human being is because of the gene, the gene is the irreducible unit, the functional unit of hereditary, the unit of life while unit of life it actually cell, but how the cell will behave what kind of proteins will form.

What kind of nucleic acids are there are all developed or all present inside the gene. Now I ask you if I change one single bit the entire message is changing, If I change the atom the entire matter is changing, if I change the gene will the organism remain as it is. Welcome to the world of genetic engineering.

(Refer Slide Time: 25:26)



Let us discuss mutation now. Mutation is the term when you have normal gene coding normal protein, but some kind of change has happened in the base pair as a result either you are getting abnormal protein created or you are getting no protein created. I will go back to my original examples of bits of 1s and 0s and we will try to explain you this. So, this was our original message that the spy received and it has asked him to come on Wednesday 4 PM.

Now, remember what I said for validity purpose it has to have a start code, it has to have a start triplet and only the information available between start and end, start and stop is valid by any chance malfunction, corruption, mistake, mischief that part gets changed. So, there is no start code in this entire stream, meaning there is no message coming up. Quite similarly if you have just one bit instead of 0 here you convert it into 1 you get a wrong message.

Instead of Wednesday 4 PM, you have Wednesday 5 PM. This is no protein formation, this is an abnormal protein formation. Formation of abnormal protein mutation creates huge amount of problems. Cancer is one prominent example as such. Though, it is much more complicated than simply having a bit changed. Again, as a homework do this, you already know from the codon chart that I gave you that what is the information.

What are the proteins that this creates? You have now just changed a base pair. See, how the amino acid sequences changes and result of which will produce in a completely different protein that completely different protein can cause more harm than good sometimes it can do good as well, but usually mutation occurs may be randomly, may be something gets triggered. There are several genes cancer causing genes an example being brca1, not Barca, not to be confused with the very famous football club all you football players might have got a heart attack.

When I read about brca1 almost pronounce it as Barca and I thought my God what are they doing. brca is a gene which causes breast cancer and brca has sets of information like this. Usually it is present in almost every human being with a varied amount of quantity, but something triggers it and this say if this is brca gene it starts produce large number of abnormal proteins.

And this large number of proteins causes the entire disease to manifest and cancer occurs. How is this switched on, how exactly is this disease forming, how exactly remember just because you have a gene does not necessary mean that it will be converted into protein. You have sent this information to the spy, the spy might not be listening or even if he had listened Wednesday 4 PM he is too lazy to go.

So, similarly the protein that has formed out of it either very low quantity or did not get switched on. So, there was no receiver at the end so it did not come out all of these factors may happen biology there will be exception. It is not physics, it is not Schrodinger wave equation you will be pinpointing something. So, what can cause a cancer gene to simply trigger, well, several external factors.

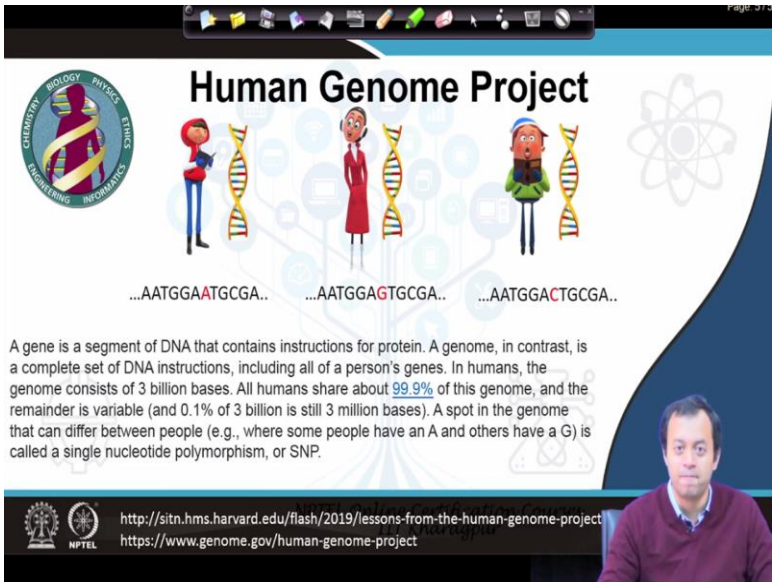
You know about smoking do not you in your cigarette packs there is written smoking causes cancer. How do you think that causes cancer? That may switch on this particular gene or that may mutate a particular gene and instead of producing normal protein it starts producing

abnormal protein or if the gene itself is that of to produce abnormal protein, cancerous protein, but it was lying dormant it has simply switched it on.

But then again all of you know somebody who has smoked a pack of cigarette since he was probably 11 years old every single day and is up till 70 years old and without any case of cancer and you know someone who has not smoke, not drank alcohol and he is having cancer. How did that happen? We do not know we are still trying to ascertain if we are able to understand maybe we will be able to go somewhere near cure.

One more question from my side. Can you change any of these gene? Can you mutate any of these using light? Remember I am focusing a bit more on genetics because the last chapters are Optogenetics where this background information will come into handy.

(Refer Slide Time: 30:52)



One such way in which we can read all the information that is present in human DNA and try to see what are the disease-causing gene, what gene makes us susceptible to say diabetes say being symptomatic to Corona virus, can attribute one such process or one such effort was the human genome project.

Human genome project started in 1990s. It was one of the biggest endeavors, intercontinental endeavor containing different laboratories both public and private from America, Japan, China, Spain, United Kingdom, France, etcetera to map the entire genome sequence of a human body.

Now, what is this genome thing? I have thus far discussed gene, what is genome? Genome is the total information. Remember in that stream of the data that I send only one part was valid between start and stop and that is gene. Genome is the overall information present everywhere. Analogy if we have a book any book the entire words written in that book, the ink, the paper basically the hardware that is DNA.

The physical material that you touch, the difference between hardware and software the paper, the ink, the illustrations, the margins etcetera, what you can touch is the DNA. The entire story starting from the name of the author, publication, foot note, the references etcetera, the entire story of the total book is the genome and gene is the climax for the most important part.

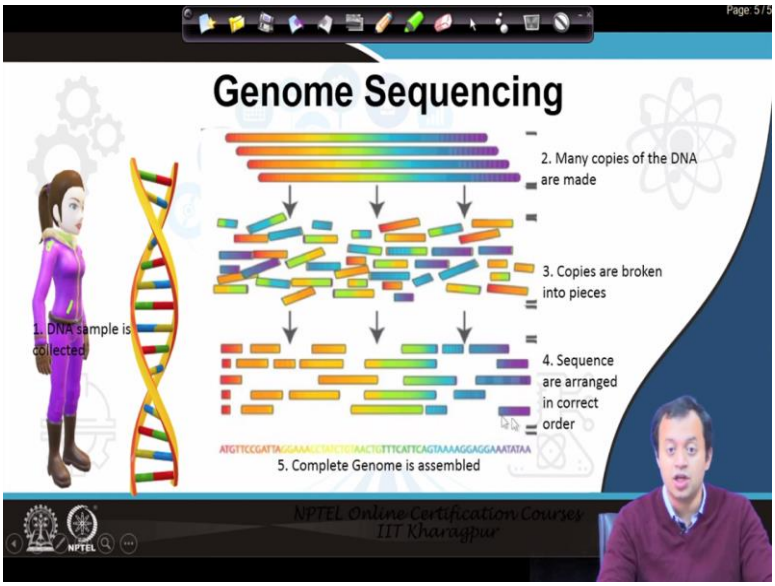
So, the physical part, the hardware part is the DNA, the software part is the genome, the story in the book and by story, I mean everything the name, the author list, acknowledgement, preface, references everything every bit of information that is carried inside the book, inside the physical thing, the paper, pen ink marks, etcetera is genome. The climax of the story, the summary, the most important part of the story, the entire novel, the entire Mahabharata, entire Ramayana, entire anything.

The main final story, the final conclusion or the gist or the abstract again these not proper and an absolute analogy, no analogy is correct, but you get it, it is gene is a subset of the genome and they started collecting information and put it in some kind of a memory device, memory chip to decide, to decode, to understand what all these information actually means and that is where we biophonotics people comes in.

We do not care much about how DNA is converting into RNA. We are interested in the final product that is the information that is present in DNA, information that is coming out of the protein. So, remember all human shares about 99.9 percent of this genome you have 3 billion bases of all the bases, all the streams that you were looking into, you have 3 billion bases and there are repetitions and only 0.1 percent of the 3 billion which is still large is variable.

This 0.1 percent of the genome that is different between you and me is what creates or what differentiates between you and me meaning you and I are 99.99 percent similar. So why fight war and go to all these troubles. This is scientifically proven that you and me are 99.9 percent similar there is no point to fight for this 0.1 percent, but anyway I digress.

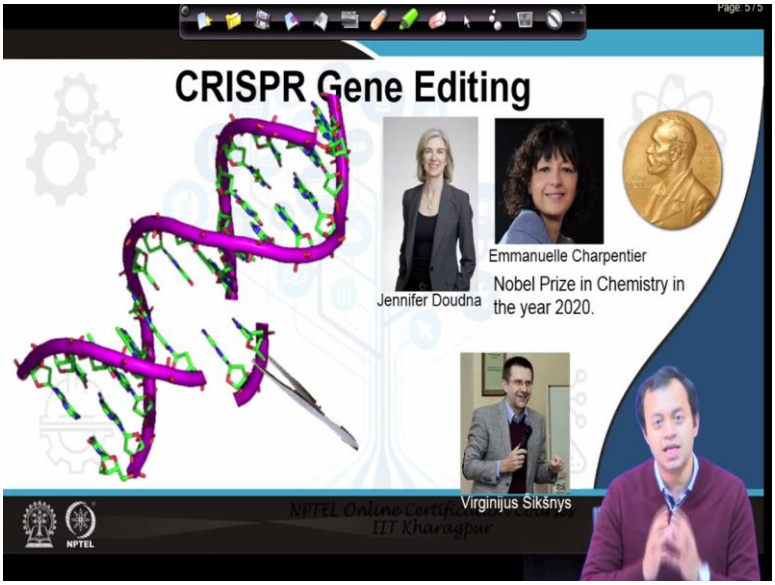
(Refer Slide Time: 35:12)



What exactly happens you take DNA sample, you make many copies, you break that down into the copies, to take out the errors, take out the repetitions, you rearrange them into jigsaw puzzle and then you read the entire stream and this goes on 3 billion times for one single human being and then we try to get it for every different representative of a human being and then we study this part of the gene is making my color of my eye look black.

This part of the gene is making me diabetic, this part of the gene is making me asymptomatic to Corona virus. This is fascinating. Genetics is fascinating and we should all learn at least, we should all make an effort to learn at least a small bit of it.

(Refer Slide Time: 36:04)



Finally, I will end today's class on this note. We have previously understood the genome assembly, we have previously read the entire book I have been given a book I have read it. Can I now edit the book? Can I now modify the book? The answer is yes recently in 2020 Jennifer Doudna and Emmanuelle Charpentier got the noble prize in chemistry you must have read in newspaper for this gene editing part.

They use this Cas9 protein this CRISPR gene editing part in which a living human being genome can be modified, can be changed. So, part of my gene is responsible for creating my color of the eye to black I am now interested to get a blue color eye. Yes, probably I can do it in near future as I said if gene makes me if I change my gene will I still be me. Here one more thing needs to be addressed and that is the case of professor Siksnys from Lithonia.

Professor Siksnys from Lithonia first came up with this idea and did experimental modification or experimental proof of CRISPR gene editing where he showed that CRISPR gene editing using a specific protein can be utilized to change the genome of a living being. He wrote this paper and sent it to think cell biology.

The cell biology without even peer reviewing simply rejected it. He then rewrote it and send it to I think PNA is proceeding of national academy of science he sent it there it took two to three months period of time. By the time when this correspondent letter was going when he was first sent nobody believe this could be done.

Jennifer Doudna from University of California, Berkeley and Emmanuelle Charpentier from Max Planck institute for I think pathogens deduction of pathogens sent the information to science within two weeks the paper got published and as a result they are getting the noble prize.

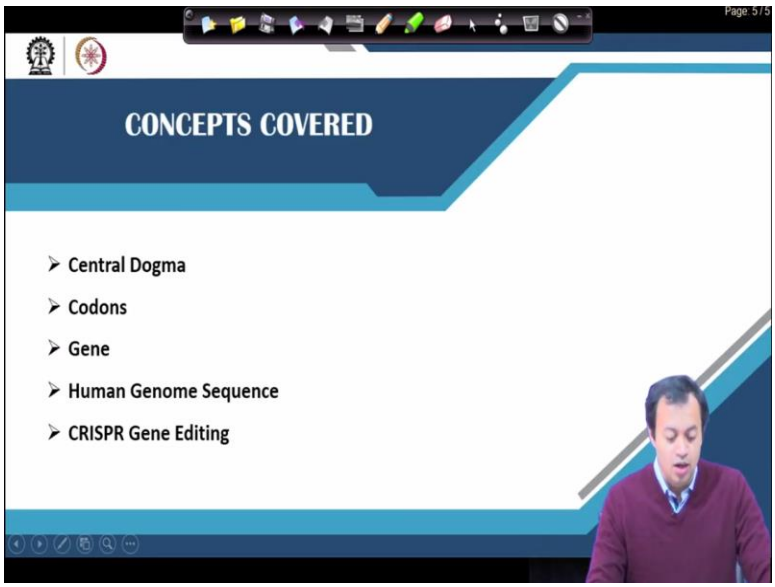
And professor Siksnys did not get noble prize. This is no way to criticize either the noble prize committee or question mark the genius of anyone of them. It is simply to say that many times it happened unfortunate as it is, but it is part of life that if your pedigree is not that good enough, if you have a made a very high scientific discovery people will not take you seriously.

The genius of Doudna and Charpentier is actually as well as professor Siksnys which has to be acknowledged at the end is to come up with this idea and then proving it using this Cas9 protein, thebacteriophage protein that you can do a genetic editing of a live person, of a live

organism. Think about the manifestation of that you do not have to kill something or you just changing a cell, taking it out and putting it inside human being.

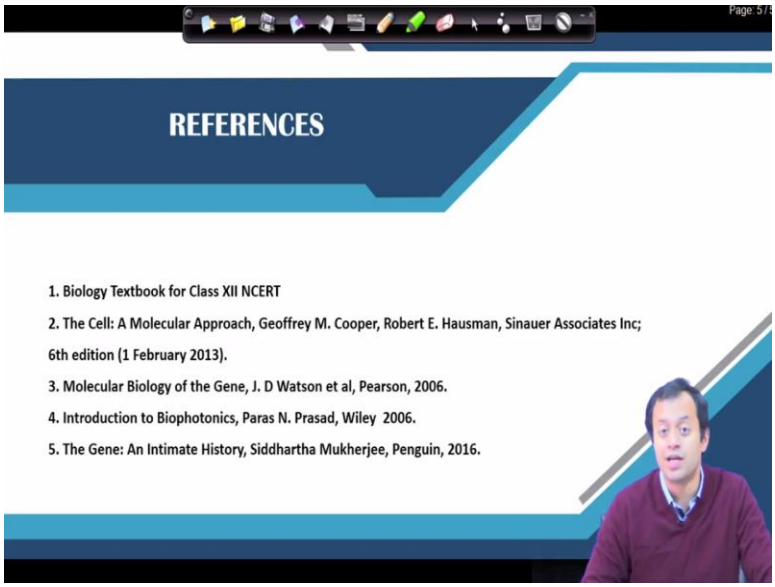
You can simply take it in a living organism and therefore got the noble prize in 2020. It is considered one of the biggest biological discoveries of the last century this CRISPR gene editing.

(Refer Slide Time: 39:58)



So, the concepts that we covered are central dogma, codon, gene, CRISPR gene editing and human genome.

(Refer Slide Time: 40:05)



And these are my references. If any of you are interested for a light reading on gene not a text book-based reading, but a popular science. I strongly suggest this book the gene and intimate history by professor Siddhartha Mukherjee from New York University I believe Siddhartha Mukherjee he got Pulitzer award as well as Padma Shri and we are quite proud of him. His book the gene an intimate history is very, very nicely written for people from any background.

You do not have to be a genetic engineer this is not a textbook. This is something for light reading and it had got huge praises from Bill Gate of Microsoft, Hugh Jackman, Wolverine etcetera, yes, I am a movie buff. So, there are two more chapters remaining, two more lectures remaining for biology basics I will finish it in the upcoming days. Thank you very much.