

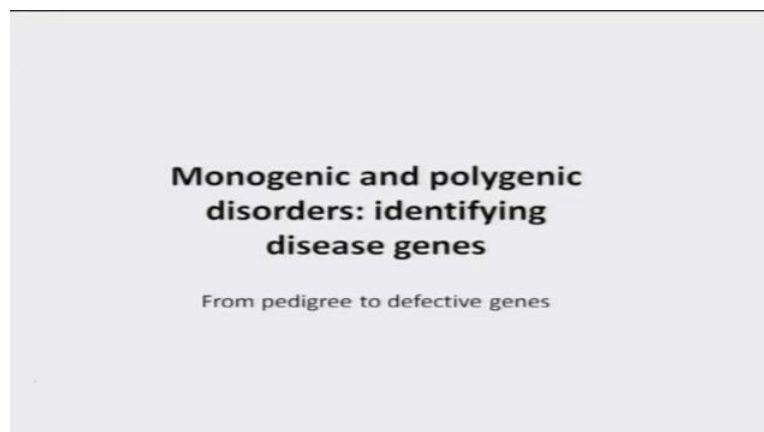
Human Molecular Genetics
Prof. S. Ganesh
Department of Biological Sciences & Bioengineering
Indian Institute of Technology, Kanpur

Module - 04
Lecture - 12
Disease gene discovery

Welcome back to the fourth week of this NPTEL course on human molecular genetics. This week also happens to be the final week, wherein we will be discussing more recent developments in the area of human molecular genetics. What you have seen in the last first, second and third week lectures is how the genetic material is, you know, sort of understood with regard to its contribution to human disorders. We looked into, for example the central dogma and how the information is processed in biological system. Then, we moved on and understood the concept behind recessive and dominant characters and variations therein and then we looked at pedigrees and looking into phenotype that segregate in the human population to arrive at a conclusion whether the genes that is causing a particular phenotype is dominant or recessive in nature and then, we looked into some of the complications with regard to the Mendelian segregation.

Then, we moved on and looked into some of the molecular biology techniques people use in characterising what are the defects that are there in the gene and how that may contribute to the disease condition. Then we discussed some of the issues about model systems and how model systems are used to understand disease pathology to validate some of the hypothesis and finally, to use it for some therapeutic applications. So that, as these are the topics that we covered so far and in this, we are going to look into the approach we use to identify the disease causing genes.

(Refer Slide Time: 02:02)



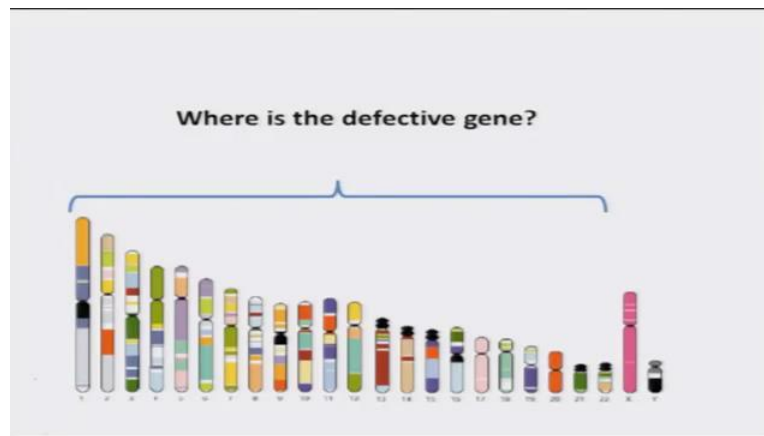
So, this is going to be a challenging topic, because it is going to have, I really, you know, expect you to have the fundamentals clear, otherwise it would be difficult for you to follow. So, even if you are having some difficulty in understanding this lecture, I would encourage you to go back and watch the, week II and week III lectures and come back and revise and, and then see whether you can understand. If you have any issues in understanding, you can always write to us and we will try to help you and of course, we are going to have hangout sections at the end of the course, wherein I could also, help you with some of the concepts, if you still are unable to understand some.

So, let us look into what is the topic that we are going to study in this week. The title is, from, you know, pedigree to defective genes. So, we have looked into the pedigree analysis, we have looked into the gene structure and how defects cause particular disease. But, what you are going to see today is how we use pedigree to identify the gene that causes the disease. So, the disorder that we look into in the human population can be categorized into two groups that have the genetic origin. One we call as a monogenic disorder, because here the, the defects lies in one particular gene, therefore you have the phenotype.

The other group is polygenic disorders, wherein the disease phenotype results from a combination of variations in more than one gene, may be 4, 5, 6 depending on what is the complexity. So, the approach that we used to identify the genes that contribute to the disease varies. If you are using a monogenic disorder to identify a disease gene, we use a very different kind of approach. If it is a polygenic disorder, we use a very different approach. So, what we are going to do is to cover these two along with some other advancements, like for example in the last two decades we have seen tremendous advancement in molecular genetics, molecular biology. The human genome has been sequenced. Now, we are trying to understand all the variations that are there in the human population. So we will, what we will do is that we will combine the approaches with how the advancement has really helped us to arrive at different approaches, right, to identify the genes. So, that would be the focus of, you know, this weeks lecture.

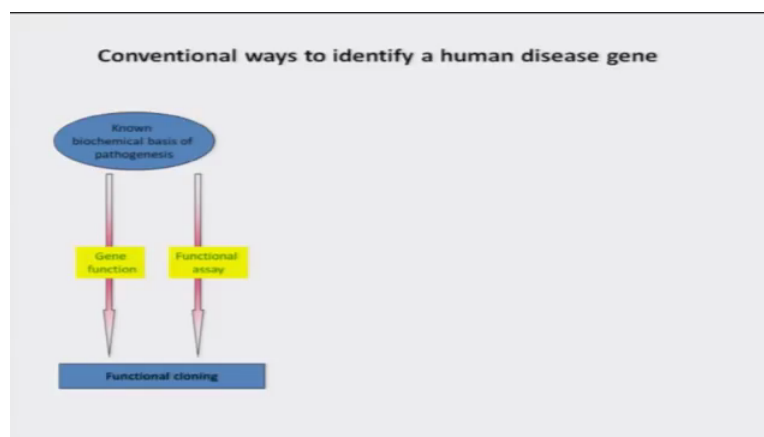
Let us look into the pedigree. So, we will start with the pedigree. So, this is something that we should be familiar with. Now, looking at it you will be able to say this is a autosomal recessive inheritance and the phenotype is because of autosomal recessive conditions.

(Refer Slide Time: 04:40)



So, we have seen all these things. The question that we are going to address today is what is the current method or how people really used what approach they have used to identify the genes? So, when you say autosomal recessive disease, we are talking about a gene that is located in one of the 22 pairs of autosomes. So, we are excluding X and Y, because they show very different kind of inheritance. Now, that is all that we can do with the pedigree analysis. You have looked at the pedigree you have, you know, you come up with the model, which suggest it could be autosomal recessive. But, now it is going to be a daunting task as to which one of the 22 chromosome that you see in this slide harbours the gene, right? So, that is going to be extremely difficult task to see.

(Refer Slide Time: 05:35)

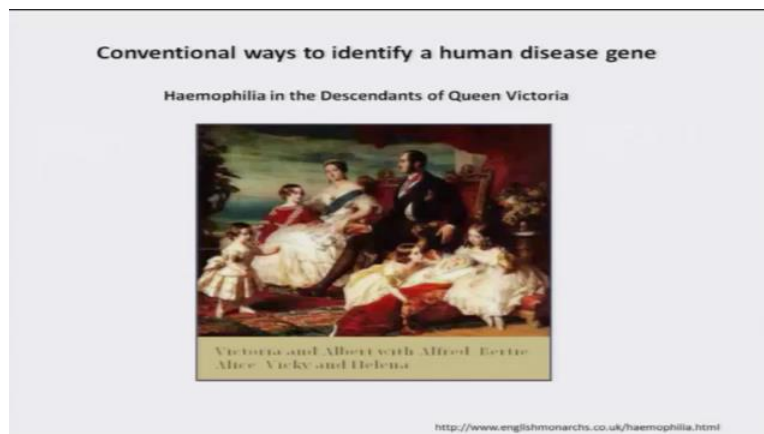


So, it is going to be a daunting task to come up with kind of a valid proof to say in which one of the 22 pairs of the autosome that we have, the gene could be located. So, what we are going to discuss is some of the conventional ways people used to identify human disease genes. So, let us look into what are the ways people have identified at least in the past, the disease genes. One of the initial discoveries on genes that are causing human genetic disorders have come from understanding the, knowing the functions of the gene or using functional assays. So, in other words, you know a

physiological pathway that is defective in a disorder and you know that disorder is caused by some genetic defect, because it runs in the family.

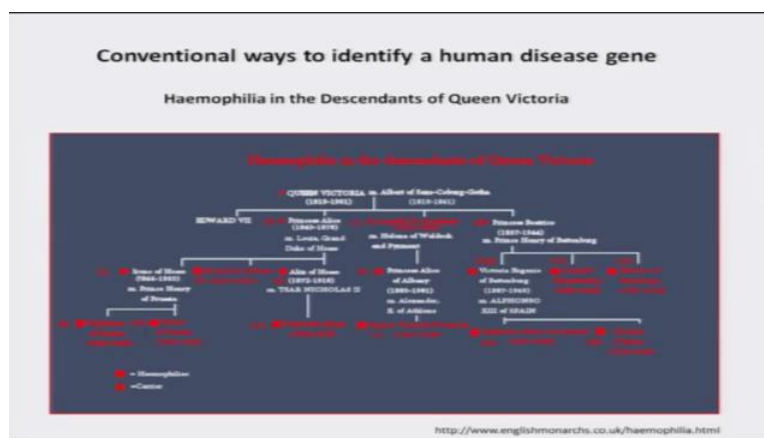
So, now you use your understanding on the physiology to identify the gene that could possibly contribute to the disease. So, this is called as a functional cloning, because you identified the gene, because of the known functions of the gene. So, that is why you have gone on to identify them. So, let us look into some of those examples.

(Refer Slide Time: 06:46)



One of the famous example is the bleeding disorder, what you call as a haemophilia that was, you know, discovered in the family of Queen Victoria, the then Queen of the British kingdom and this is one of those rare photographs that you can find in the website about the family. This is a very happy family; at least in the picture it looks so.

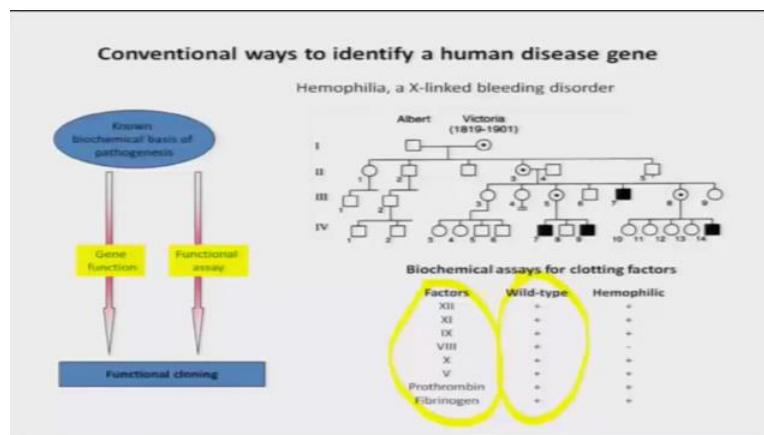
(Refer Slide Time: 06:58)



But, there are many unfortunate incidences in this family. So, as you can see here, this is a kind of pedigree that is shown and you have some individuals that are identified with a red colour filled in

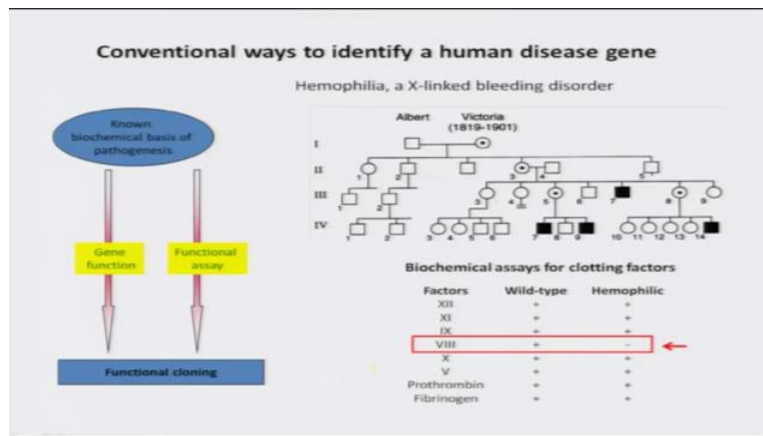
symbols for male and female. These are the people who are affected with this bleeding disorder. It has a very very severe condition, because in this condition the blood will not clot. As a result, you know, unless or otherwise you are very careful, we are going to have extremely, you know, high volume of blood being lost, because of some injury or cut or anything. So, you can see here that the Queen herself was not affected. She was all right, but she is expected to be a carrier and it is sort of perceived that the mutant gene could have come from her father and you can see in her next generation, one of her sons, he was affected and we know now that this particular disorder is caused by a gene that is located on the X chromosome. So it is X-linked inheritance.

(Refer Slide Time: 08:01)



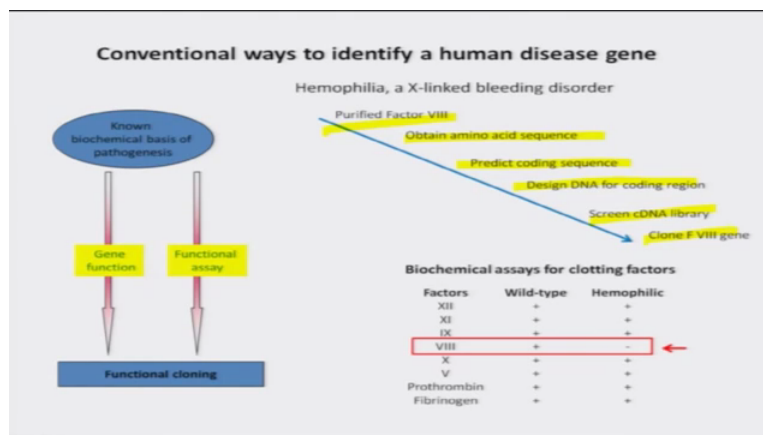
So, let us look in how they have really identified. So, this is a kind of symbol that I have shown, but let us not worry about it. What is, what has been understood is that by then people have understood the biochemical basis of how the blood clot happens. So, it is known that we have a large number of factors, proteins that are there in your blood and which, you know, upon activation, for example when it gets in contact with the air, oxygen or atmosphere, because of an injury or whatever, then it sets in a cascade which results in the blood clotting. So, there are many different factors that are listed here on the right side, the left side of the screen, like you can see here. So, you can see that these are the factors that are involved in the blood clotting. What you find in all of us who are very normal that you have all these factors are very active. So, if we do the biochemical assay for these factors, you will find that it is working.

(Refer Slide Time: 09:14)



So, what they have done is, very similar kind of approach they have used to identify which particular factor could possibly be abnormal in this family, you know, those who have the haemophilia and they found was this particular factor, at least in this family, that factor was defective. In these individuals, you do not see factor VIII being active. So, that is how they understood probably this is what is deficient. It could be some genetic mutation that results in the deficiency of factor VIII.

(Refer Slide Time: 09:30)

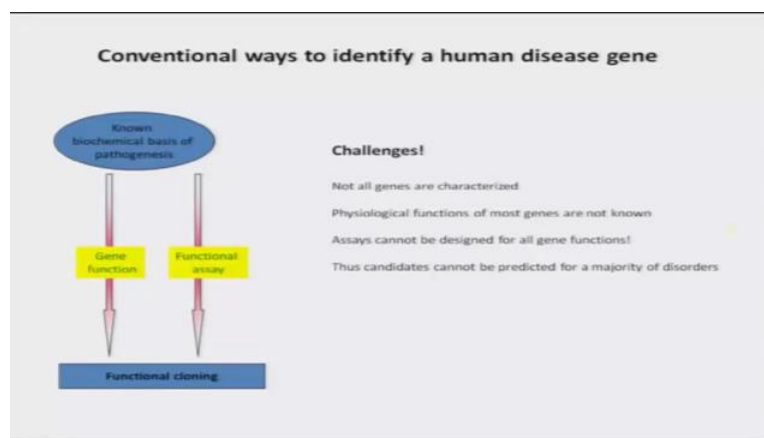


So, with that understanding they went on to identify the gene. So, the approach that they used then was that they purified the factor from the blood. So, it is, when you say factor it is a protein and then they sequenced this protein to obtain what is called as amino acid sequence, right. So, they have now the amino acid sequence for the protein and then they predicted what could be the coding sequence. So, you know what are the different codons for each amino acid and with that you can come up with multiple options as to what would be the coding sequence and based on that sequence, they have designed DNA, short stretch of DNA using lab machines. Now we have what is called as a DNA synthesizer. We can add those, it is a kind of chemistry by which you will be able

to add certain bases to get kind of sequence that you want and they have designed this coding region and then they went and screened what is called as a cDNA library.

Remember, we discussed about that that we convert the messenger RNA into a single stranded DNA and then to double stranded DNA, we clone them in the vectors and then they are stored and you can use any kind of a probe what you call here for example, the DNA coding region for factor VIII,. you can use it as a probe and identify the cDNA that codes for this factor. That is how they cloned gene and then they went on to say what is the defect in that particular individual. So, that is how they have done.

(Refer Slide Time: 10:57)



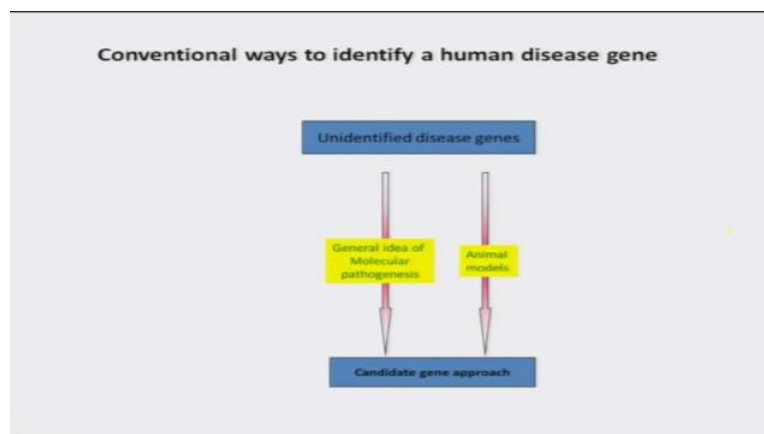
So, this approach that led to the discovery of the gene is really really is time consuming and you have lots of challenges. I know there are few such examples, wherein people identified the protein product that is defective and then went on to identify the, the gene that codes for that protein. So, there are challenges. Some of them are listed in the screen. Not all genes are categorized. In fact, we do not know all the genes that our human genome contains and even if you know the gene, we do not know what is the function of the gene. So, you may know this is the protein that it codes for; that is all you can predict, but still you do not know what is the function of this protein in the organism. For example, you may say it is an enzyme present in a cell, but how does it really, you know, change the individual, right, that is not known.

So, the physiological functions of most of the genes are not known and therefore, since you do not know the function of the gene or the protein, you cannot, come up with assays to test whether it is active or inactive, how mutations could have affected and so on. A protein could have multiple different functions and how are you going to come up with all sorts of assays? For example, you are talking about their mental retardation as a condition, right? So, wherein your intellectual ability is

not that good, your IQ is low. So, what kind of assays you will do to understand how the gene defect affected your intellectual abilities. It is going to be extremely difficult. So, even designing assays for some of the genes of some of the phenotype that you are looking at are going to be extremely difficult. So, therefore even predicting, I mean these are the candidate genes, therefore I can go and identify mutation, this is also, is difficult, but still there are some examples.

For example, epilepsy, one of the conditions wherein people get what you call as fits. You know, they thought that it could be resulting from abnormal neuronal function. Therefore, people thought it is the channels that regulate the, what you call as the neuronal functions in terms of its membrane potential, how the electricity is passed on from one to the other, these are regulated by channels and people thought the channel could be involved and they went ahead and sequenced many of the channels. There were some success, they were able to find some mutation, but still it is extremely difficult approach to identify the genes based on the functions of the protein that you know. So, therefore it is very very, the success has been, very very limited.

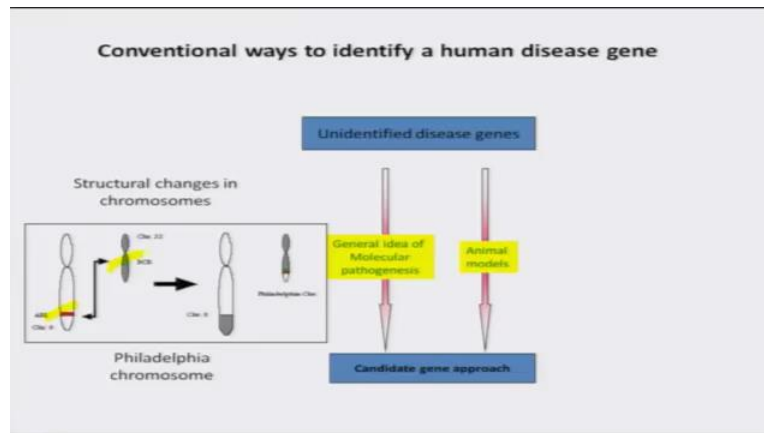
(Refer Slide Time: 13:35)



The other approach people have used is to look at animal models, you know, when you talk about disorders, disorders are caused by defect in a gene. So, we have genes, likewise every living species have got the genes and defect there in should also result in some abnormality. But more often what we do is, we breed domesticated animals. We breed them and we have selected for some good characters. If the character that comes in a generation that are not desirable, more often that are not, you know, selected by the human. Therefore, the gene gets lost, but still there are many condition that you see wherein, there are animals that developed some symptoms, some phenotype that resemble some human disorders and when they have identified such, then they have located the gene for that using human, human, the sorry, using the mouse species or any other model that they have used and once you know this is the gene that causes a phenotype in this animal, then you can

go and look at the mutations in that equivalent gene in the human. So, this is also has been some of the approaches.

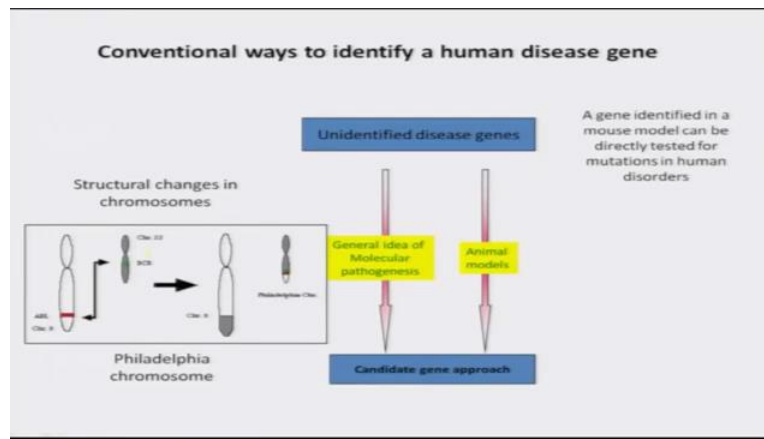
(Refer Slide Time: 14:43)



I will show you one such example wherein, so this is a kind of approach people have used looking at the function of the gene product. See, the other approach people have used to identify a gene is based on structural changes in the chromosome. So, as we discussed, the chromosomes, at times there are exchange between the chromosomes. As a result, you have rearrangements and the rearrangements are known to result in certain disorders, right? So, therefore if a given translocation and what I mean is that a part of the chromosome goes and attaches to some other chromosome, you have a disease. If you are getting the same disease and the translocation is seen in many individual because of recurrent events, then you would assume that the gene involved in the onset of the given disease is located very close to where the chromosome is broken, right? So, that is the approach, so they have used and there are many examples.

This is something I am showing that we discussed already that is the Philadelphia chromosome associating with a type of blood cancer is resulting from translocation between chromosome 9 and chromosome 22 and we know that the break point where the chromosome break happens, here and here, you know, harbours genes ABL and BCR and because of the fusion, you have a new sequence and one gene gets activated, it would normally should not be expressed and then you have the disease. So now you know, this is in the, for example ABL or BCR is involved in, you know, this particular cancer. So, this is another way by which, you know, by looking at chromosomal changes, translocations, people have, they are able to identify the disease causing genes, right? So, again later they have understood what is the function; so this is the, one of the ways by which we can do.

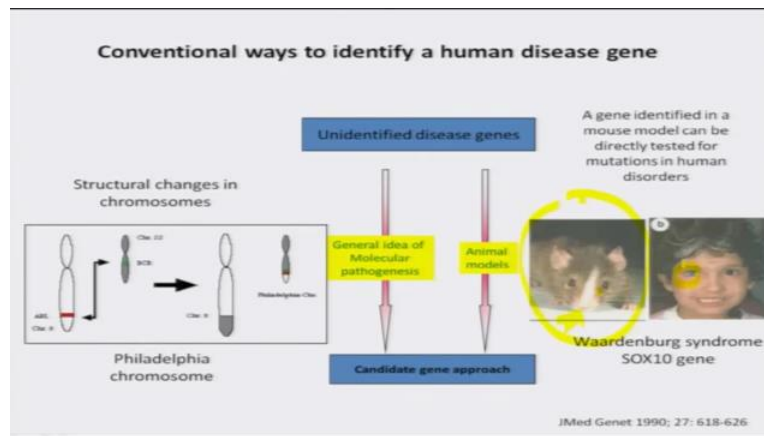
(Refer Slide Time: 16:35)



So, the third, approach people have used is to characterize animal models. There are animals that we use for studying genetics and many of these animals also have, genes that are very, very similar to us. For example, if you look into the mammalian species, whether it is a rat, mouse or any other model system that you study have got all the genes that we also have. Therefore, if there are some defect that we expect to cause a genetic condition in humans, is likely that the same gene, gene defect in the mouse or rat and rabbit should also be causing this similar phenotype. That we have discussed already in the gene knockout lecture, wherein you are able to create deletions, create mutations in the gene that cause disorder in the human.

So, you create similar mutations in the animal and then look at whether the animal develops and if it develops, then you understand what is the pathology. But people also looked at animals, which some how inherited a natural mutation. As a result, they have certain defect and resulting in a phenotype and if the phenotype resembles some of the disorder that you see in humans, then you know that the gene that is mutated in mouse is also likely to be mutated in the humans. So, you identify what is the gene that is defective in mouse and look at the corresponding gene in the human and see whether that has defect.

(Refer Slide Time: 18:09)

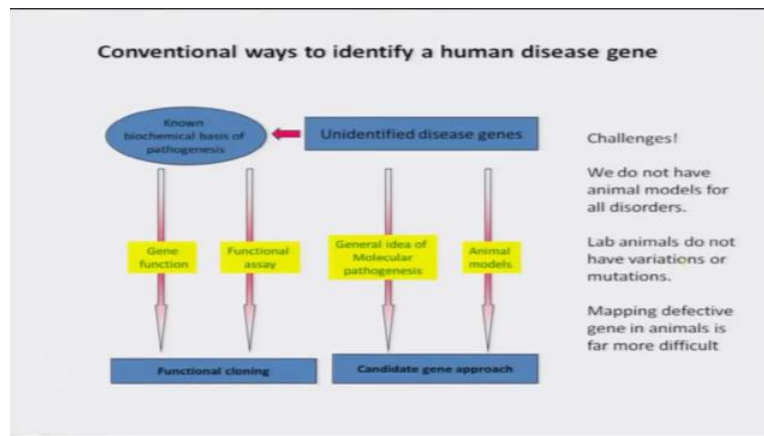


There are some success stories there too. What I am showing is a condition called as Waardenburg syndrome and this particular syndrome is caused by a gene called SOX10. This is a transcription factor. So, what is this disease? This is a congenital anomaly, there are multiple developmental problems. One of the phenotype is also something to do with pigmentation, whether, the colourations. So, what you can see on the top is that this particular mouse, you know, this naturally developed a mutation resulting in many developmental abnormality, as well as pigmentation you can see here in the fore head that the hair colour, it is very very unique, you do not have the colouration and also the other condition is the distance between the eyes, between the two eyes also is enlarged. We have more flat face and that there are other problems, there are a number of problems.

What is interesting is that very similar phenotype has also been seen in human. For example, you can see here the baby; she has got premature grey, you know, grey hairs and then there are problems with eye and there are problem with, you know, heart development and many, many different conditions. What is , you know, important to understand from this particular course point of view is that they identified, they did a mapping for the mouse gene as to what is the gene that causes this phenotype in the mouse and they identified the gene called as SOX10 and then they looked at that there is a deletion that it causes and so on. Once they understood the gene, then immediately what they did? They did not do any mapping for the humans, straight away they looked at whether SOX10 is a gene that is mutated and indeed that was the case. You know, they went ahead and then shown that it is similar gene mutation, similar phenotype in the humans as well.

So, there are handful of examples wherein animal models which developed, one or the other this condition, because of spontaneous natural mutation led to the discovery of genes that causes the disease in the human as well. Again, these are limited. The challenges, we do not have animal models for all the disorders.

(Refer Slide Time: 20:36)

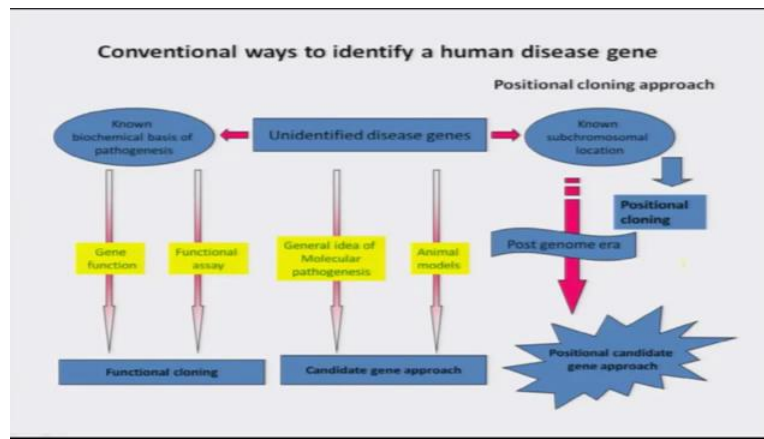


That is, it is not possible, because of reason that most of the model system that we use for our experimental purpose are in built and maintained in the colony. Normally, look at whether the animals are normal, they have a healthy phenotype, if not we remove them. Therefore, we know that whatever experiments we do, whatever effect you see is because of the experimental change, but not because of something, not linked to the experiment that you do. Therefore, you normally maintain very healthy animals in their lab settings. Therefore, you will not find any natural mutations coming in. Rarely you see that and lab animals do not have variations or mutation like, what you see in human.

All of us look different, all of us behave different and all of us are exposed to very different kind of environment, whereas lab animals are in bred, meaning they have very very similar genetic make up; variations is very very limited and then, therefore you keep selecting for a, you know, the good phenotype. Therefore, you do not really allow them to sort of acquire in the population, some abnormal mutation or variation in the DNA and even if there were mutations resulting in the phenotype in the lab animals, mapping them is going to be extremely difficult. It is more difficult than what you can do with humans, because normally you do mapping based on the sequence variation that you have in the genome. That is what we are going to discuss now.

But you know, the sequence variations if you, look into the lab animals, they are very, very limited, that is what I was explaining that they are what is called as isogenic line, meaning you minimize all the variations at the genetic level, at the environmental level and only you look at changes you bring in, because of your experimental condition and the phenotype that you see in.

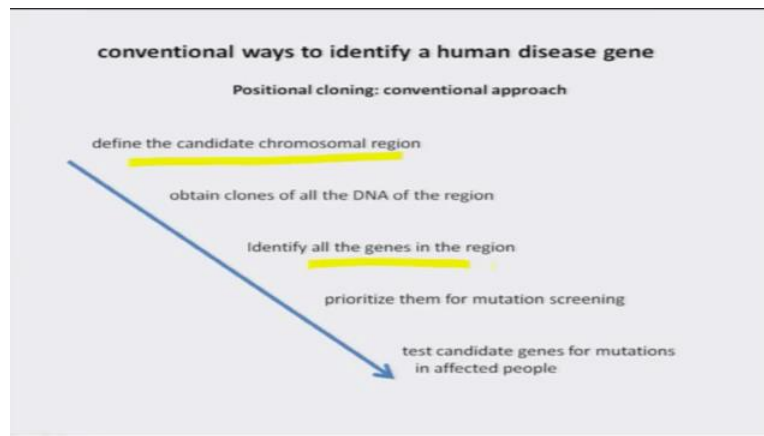
(Refer Slide Time: 22:20)



Therefore, it is extremely difficult to use, model animals to identify human genetic, or, or the defective genes for human disorders. So, what is the approach that was very successfully used to identify a majority of the genes that cause monogenic disorders? The approach that people used very successfully, thousands of genes have been identified in the last two decades is called as positional cloning approach, meaning you look at region of a chromosome that is associated with the disease.

Without really bothering about what is a gene, what is the function, what is the pathophysiology, you only look at a region of chromosome that is likely to be contributing to the disease phenotype and then look at what are the genes that are there and then look at what variations there results in the disease condition. So, that is called as, you know, positional cloning approach and there are two different approaches. In the post genomic era, where we now know the human genome sequence, the approach now you use is somewhat different as compared to the approaches that were used in the, when we were not knowing the human genome sequence. So, we will talk about, the approaches that were used before the human genome sequence was understood and then how that is really not necessary now, once you know the human genome sequence, so that is what we are going to discuss.

(Refer Slide Time: 23:45)



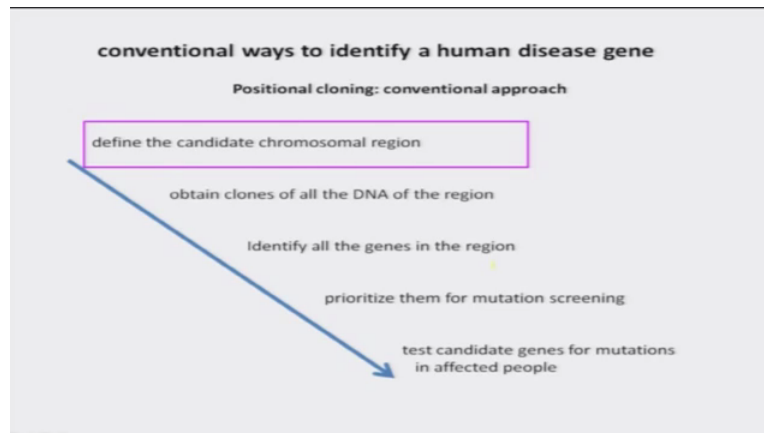
So, these are the different steps in identifying the defective gene using the approach what you called as positional cloning. We call it as conventional approach, because this is the approach that was used before the human genome sequence was made available. So, what you do? You define the candidate chromosomal region, you sort of say that is the region of the chromosome, a particular chromosome that is causing a particular disease likely and then up time get the genomic segments representing that chromosome and then look into just just block it here.

So, let us look into the conventional ways to identify the human disease genes. So, what you call as a positional cloning approach, wherein we identify a region of a chromosome that is associating with a particular disease, then go on to identify the gene that are located in that region. So, what are the steps? So, the first step is, you have to identify the candidate chromosomal region that is something that is shown here, right? So this is called as identify; so, identify the chromosomal region that is associating with the disease. That is the very first step. Once you identified, then you have to get, DNA fragments representing that region of the chromosome, you go to the genomic DNA library and identify all the fragment that represents the region of the chromosome and then you have to go and next look into genes that are located in that region where you thought, that defective gene could be present and then based on some features like, for example whether the gene is expressed in the tissue that is normally affected in the disease condition.

You prioritize your gene list and look for variations or mutation in the gene and with this approach by repeating one gene after the other you will be able to identify possibly the gene that causes the disease condition. So, again you will be looking at some variations and you will be assuming that this is the variation that causes the disease and eventually what you have to do is you have to validate. You have to use some animal models or some other systems to validate that the changes that you have seen indeed, affect the way the cell functions, tissue functions and organism

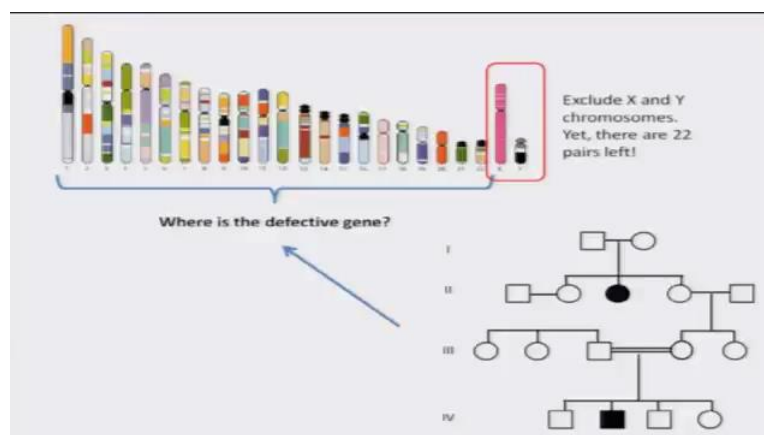
functions. So this, as you can see is a laborious process. It takes years to identify that defective gene and to characterize using different model systems.

(Refer Slide Time: 26:16)



So, first let us look into this particular aspect that how do you define the candidate chromosomal region, because the rest of them we have already discussed. We have discussed how you get the genes from cDNA library, we discussed how you do mutation screening, we discussed how you create animal models for a given gene. So, let us look into the first two topics that is define the candidate chromosomal region and then up time the DNA fragment representing that particular region and then the rest of them you would be able to follow.

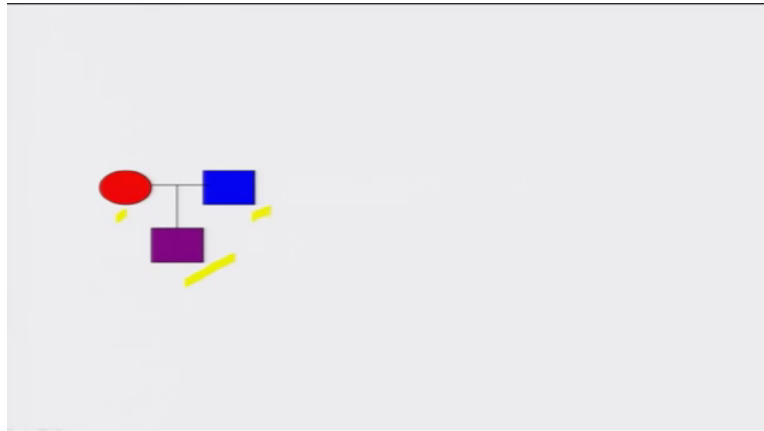
(Refer Slide Time: 26:40)



Let us look into the pedigree that is shown on the lower side of the, the lower panel. This is an autosomal recessive disorder. You can see that affected persons parents are not affected and you have consanguinity, there is a affected person in the previous generation. Because it is autosomal recessive, we know that the disorder, the gene that is defective should be in the autosome, the point that we have already discussed and we have excluded X and Y. So, how will you really identify

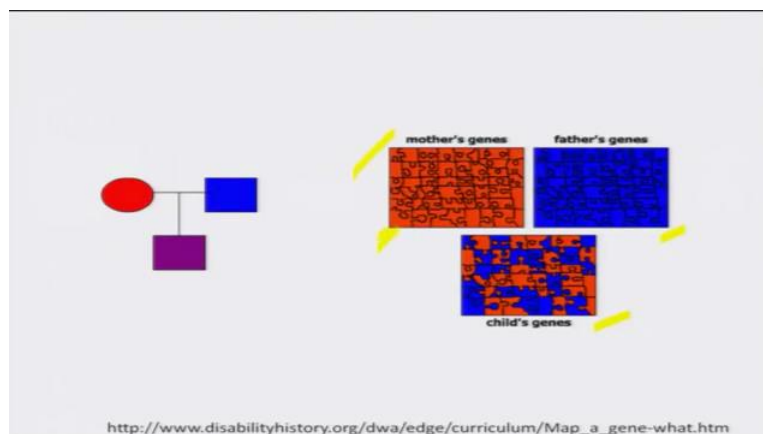
which chromosome and then which part of that particular chromosome possibly has got the defective gene. So, let us look in how we people do?

(Refer Slide Time: 27:19)



This is what we sort of, we are going to look at some puzzle that would try to sort of help you to understand the concepts. So, what we are trying to say is that the, right, so what we are going to say is that this particular individual who is a, you know, individual that are born, that is born to the male and female, he has got mixture of the DNA that he inherited from his father and mother. So, you have, 50% of your genome is derived from your father and 50% of the genome is derived from mother. To represent that we are showing it as a mixture of these two colours, but it is not necessary that these two are exclusive, I mean is not that chromosome 1 has come from father and chromosome 2 has come from mother. For every chromosome we have two copies one that has come from father, one that has come from mother. So, this is what we expect, right? So, let us look into a puzzle.

(Refer Slide Time: 28:16)



This puzzle is, you know, you have, you must have played this kind of puzzles during your childhood days. So, you have a identical piece, but they differ in the colour. For example, mother's puzzle pieces are red in colour, father's puzzle pieces are blue in colour. So, what you are expecting is that there is a mixture of the mother and father DNA and that is what you, you know, in the next generation; how will you, this is the concept. So, what we are talking about, each puzzle for example we can say it is a gene and the colour represent, for example the, let us say the dominant allele, right, for that particular piece, for example this piece here that you have seen, it could be a colour represent a different allele of a gene, let us say that way, right?

(Refer Slide Time: 29:13)

Find the gene for this disorder

The correct puzzle piece must be present in every individual affected with the disorder.
 The correct puzzle piece must be absent from every unaffected individual.

http://www.disabilityhistory.org/dwa/edge/curriculum/Map_a_gene-what.htm

Let us go on to look into, so how you have, you know, you can use this information to identify the gene. This is a kind of a, you know, a kind of a test, how quickly you can identify a puzzle piece and colour. Let us see this pedigree and this pedigree obviously represent a dominant disorder, because in every generation you find individuals are affected, transmitted by either a male or female and, and both male and female are affected, right? So, how do you really find? So, assuming that what is shown here is the genome, right and all the, each one of the piece are the different genes that the genome has, you need to identify which piece of this particular puzzle could possibly be carrying the gene that is defective and causing the disease

So, how do you go about doing it? The correct puzzle piece must be present in every individual affected with the disorder, the correct puzzle piece must be absent in every unaffected individual, because you are talking about a dominant disorder and we are assuming here it is a 100% penetrant, you would expect that if I carry the mutant allele, I would definitely show the symptoms, the phenotype. So, if I am not showing the phenotype that means I am not carrying the defective gene.

(Refer Slide Time: 30:41)

Find the gene for this disorder

The slide shows a pedigree chart with 12 numbered individuals. Affected individuals are 1, 3, 4, 8, 9, 11, and 12. Unaffected individuals are 2, 5, 6, 7, 10, and 13. A 12x12 grid of puzzle pieces is shown below the pedigree. The correct puzzle piece is highlighted in red in the original image. To the right, a 3x3 grid of puzzle pieces is shown, with the top-left (red) and top-middle (green) pieces marked with yellow checkmarks. Below the grid, two checkboxes are present: the first is checked and the second is unchecked.

- The correct puzzle piece must be present in every individual affected with the disorder.
- The correct puzzle piece must be absent from every unaffected individual.

http://www.disabilityhistory.org/dwa/edge/curriculum/Map_a_gene-what.htm

So you have to use this logic to identify which is the puzzle piece. So, this is the genotype that is given, the genotype given for each individual that are numbered here. So, 1,2, 3 and so on. So, you have father, mother that are identified here and then you can look into how the whole thing segregates

So, if we can spend time and look at this slide, probably you will be able to identify which is the puzzle piece. Again you have to go back and look at that the individual who is affected should have, all the individuals that are affected in the family must have a particular puzzle piece with particular colour that is the gene and the mutation therein, because that piece is going to be there in every individual, because you, all of us have the gene, but only a some of us will have mutation and therefore you will have the phenotype and all the unaffected individual should not be carrying that particular colour of that particular piece. So, if you look in this and you will be able to identify. So I have given the link for this particular site. This is something that I have downloaded from a popular teaching tool that is available on the net. So this is, sort of helps you to understand how we can identify the disease gene.

(Refer Slide Time: 32:03)

Find the gene for this disorder

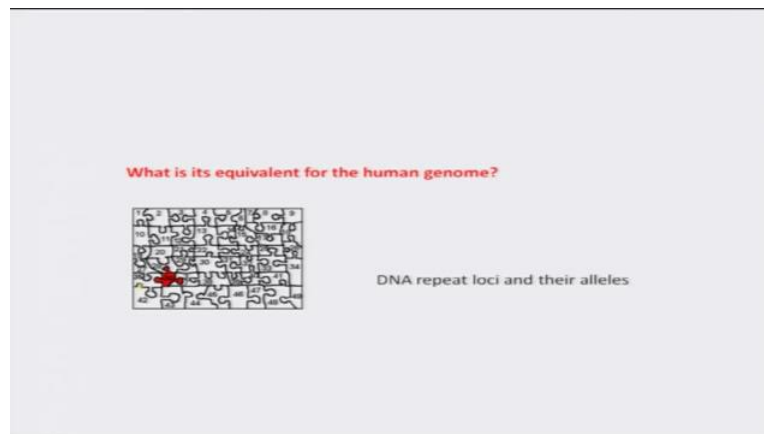
This slide is identical to the previous one, but the correct puzzle piece in the 12x12 grid is highlighted in red and labeled "Answer!". The 3x3 grid of puzzle pieces to the right remains the same.

- The correct puzzle piece must be present in every individual affected with the disorder.
- The correct puzzle piece must be absent from every unaffected individual.

http://www.disabilityhistory.org/dwa/edge/curriculum/Map_a_gene-what.htm

So, if you really go through this link and you spend time, you will be able to identify that piece number 36 and colour red should be the one that is causing the disease, right, because this is absent in every individual that is not having the disease and invariably present in individuals that are having the disease, this particular piece and red colour. So, you can go and verify by looking into this site. So that is the approach indeed that we used to identify that, you know, the disease gene.

(Refer Slide Time: 32:28)



So, what is the equivalent of that, you know, the p's and the colour in the human genome? These are nothing but the repeat loci. So you have in, your genome, in large number of the regions are having repeat sequences and these repeat sequences vary from one individual to the other and we use this as the piece as well as the colour and that is how we identify the disease gene. Let us look at how we use this.

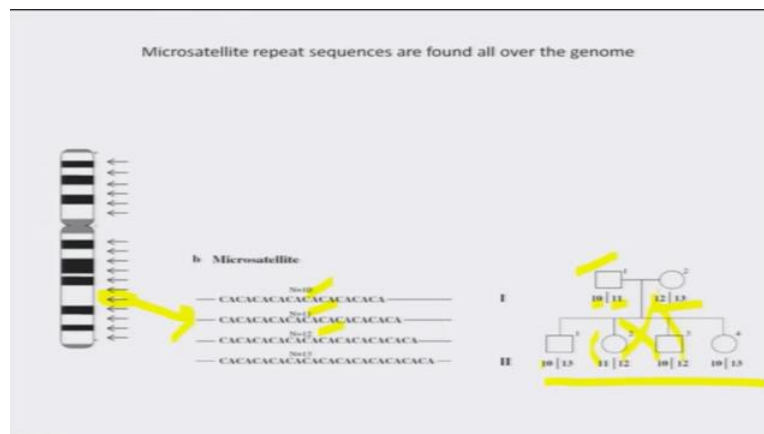
(Refer Slide Time: 33:01)



See, in all our chromosomes we have a large number of repeats, as I said. So, these are called as microsatellite repeats. They vary in size and, and how the variation in the population, for example it could have just a CA repeat, repeated many a times. So, for example, it could be 7 times repeated,

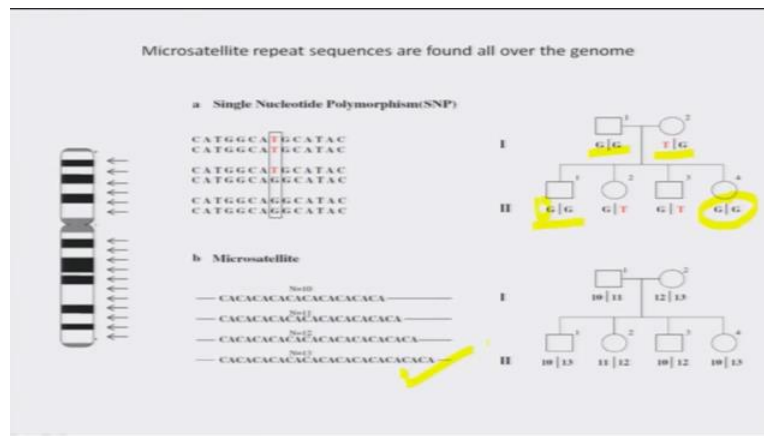
27 times repeated, 8 times repeated, varies in the population. It could be GATA, it could be GTC, it could be TA and you can see that from this telomere to that top telomere, every chromosome has got these repeats spread over the chromosome, right and this is what we use as a marker to track the segregation across generation. So, you can really say which part your chromosome has come from which grand parents. That means, you know, your father's father or your father's mother who contributed that small piece, although it came through mother. So, we can really really look at and then decipher that.

(Refer Slide Time: 34:05)



How do you do this is something that is shown here. So, you have repeat loci. We are looking at a particular region, particular, let us for example, you are looking that region, let us say. So, you have variations. In certain individuals you may have an allele which is 10, meaning 10 times repeated CA, in some it is 11, some it is 12, it varies depending on you may be homozygous or heterozygous. You may, at the maximum, for a given loci, you may have two alleles. So, if you look into a pedigree that is shown on this side, what you see is that this individual is a carrier or carries two alleles for this gene carrying 10 and 11, these are the repeats. So, a particular marker we are talking about, not the gene., This marker has got 10 repeat, the other homologue has got 11 repeat, this individual as got 12 and 13. Now, we can see in the next generation that there is a combination, for example from father 10 has come, from mother 13 has come. Here, the other allele, from father 11 had come, 12 had come from mother, 10 had come from father, 12 had come from mother.

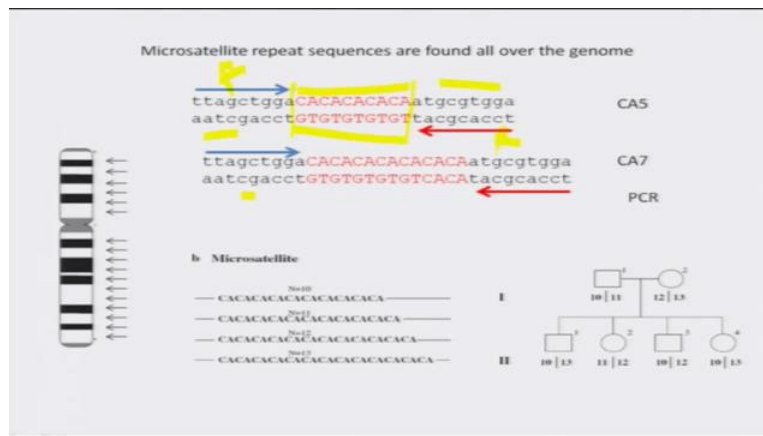
(Refer Slide Time: 35:23)



So, you can really track the segregation of each of the alleles from the parents, right and this is vary, although it is very similar to the mutation that you have seen. So, we have looked at point mutation and we have looked at how it segregates in the population, you know in a family. The difference is for a region of a gene, where your particular base that is mutated, at the maximum you may have, four different alleles that is possible, one of the four bases. So, but more often you have only two alleles, the wild type and then mutant. You do not see many more, but when you are looking at microsatellite markers like what is shown here, the CA repeat for example, there may be 20 or 30 or 40 different alleles in the population. So, that makes it much more powerful to identify which region has come from mother and father.

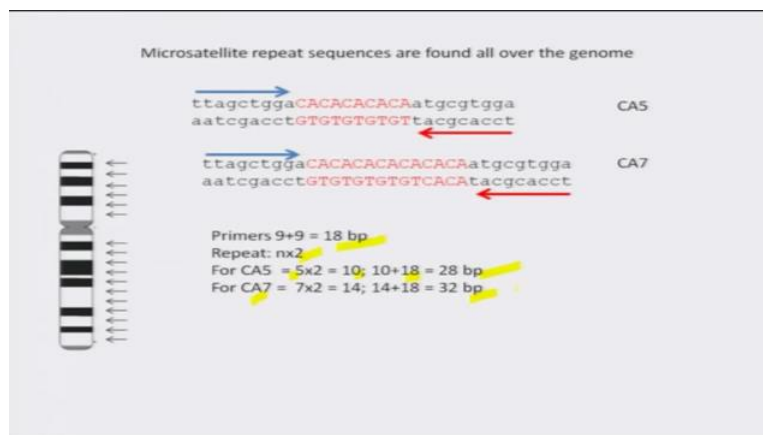
For example, if you have only two alleles like what you see here GG and mother is TG, heterozygous and you have GG. Now, you have many ways to tell which G had come from which parent, whether it is from father or it is from mother, we cannot say. Likewise for this individual, so it becomes extremely difficult. That is where the microsatellite markers are powerful, because they have a large number and the, the probability that you would have same allele in both the parents is less as compared to this kind of mutant allele that you see, where one base is changed with the other. So, how do you really type them, how do you see this segregation?

(Refer Slide Time: 36:53)



So, this is something that I have shown you that you have the repeats. In a repeats that you see here, it is CA repeat. What is shown on the top strand, this is a complimentary strand and you have, you call this allele as CA5, because it has got 5 times repeated, you have another allele that is 7 times repeated. So, how do you really type? Therefore, I can tell this is the one that had come here. So, how do you do that, you know, typing? If we do it by the conventional PCR, that is called as polymerase chain reaction, routine PCR, wherein we design two primers, one is forward and the reverse. Now, the primers as you see, have come from the flanking sequence. You have the repeat here, but this is what you call as the flanking sequence. These are identical whether it is 5 repeat or whether it has got 7 repeat, the flanking sequence are identical. Therefore, I can use the same set of primer to amplify any number of alleles, it becomes very very easy; that is one particular point.

(Refer Slide Time: 37:53)

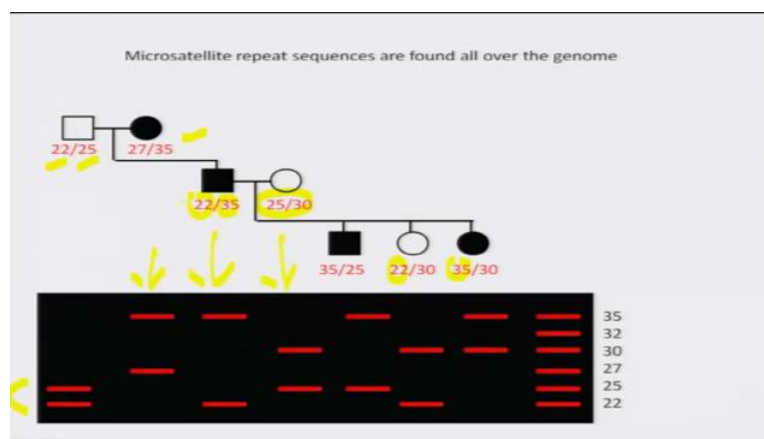


So, let us look into how you type the allele, which allele is the one that you are, you know, looking at? So, when you do a PCR, you will be able to see a difference for each allele, the reason being your primers are coming from the flanking sequence, unique sequence. So, you can calculate what is the size of your amplified product and then you will be able to tell which allele. For example the two primers that you see the two primers have a definite length that is 9 bases. Therefore, when you

add them, it becomes 18 bases and then you have the repeats. The repeats are dinucleotide repeats. You have two nucleotide unit that is repeating CA, CA and so on. So, how many times it is repeated that would tell you what is the length. For example, if you have 5 repeat, then it is going to be 10, so plus if you add 10, plus 18, which is this sequence of the primer, now you are going further 5 allele it is going to be 28 bases; for 7 allele for example, it is going to be 32 bases.

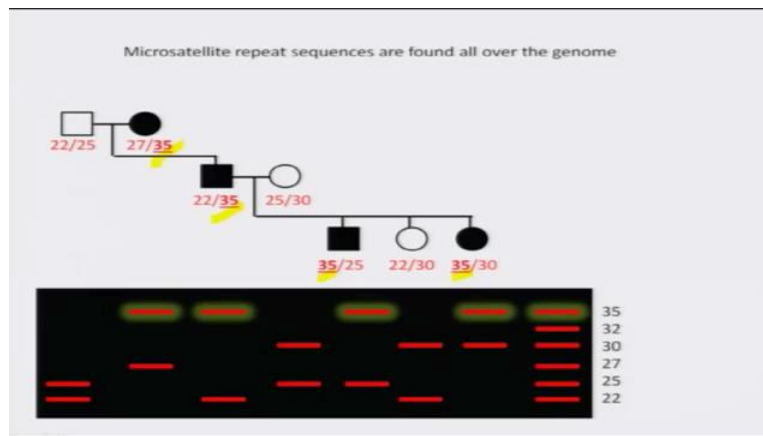
By looking at the size difference in a gel and calculating what is the size of that particular fragment, we will be able to tell what is the allele that you are looking at. So, this is the approach people use and when you, obviously when you run these products in a gel you are going to find distinct fragment, DNA fragments you can expect the size and then calculate.

(Refer Slide Time: 39:14)



So here is a pedigree which shows an autosomal dominant disorder. We can see that individuals in every generation are affected and what you have done is, you have looked at a marker, a marker by PCR and what is shown here is the, how it migrated in a gel. So, each of the numbers that you are showing, either below an individual or in the gel represent the particular allele; 25 is an allele, 22 is an allele and so on, so forth. So, you are looking at an autosomal marker. Therefore, individuals would be having two copies that is what you are seeing here. This individual in the first generation is having 22 and 25 as you can see here and she has got 27 and 35, again that is shown here. Likewise, you show for every individual what is the genotype that you see. So, basically you look at the gel and then you derive what is the allele that is got. Now, you can clearly map it. For example, this 35 has come from this individual and this 22 has come from this individual and so on, you can map.

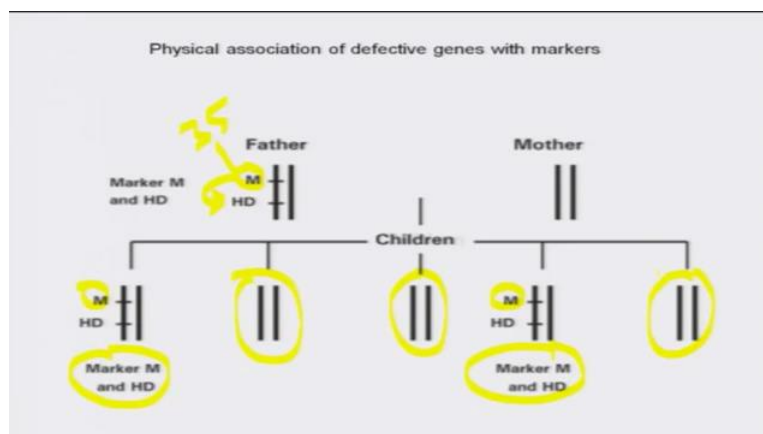
(Refer Slide Time: 40:21)



Now, how do they really help you to identify the gene? So, you can see there are certain allele markers, now that are highlighted and these are the markers that are invariably present in the affected individual. If you can go back and connect with the puzzle that you have seen, so you have looked at a piece and you have looked at a colour. So, what you have looked at that particular piece having a particular colour should invariably present in the affected that means that is the piece that possible is likely to have the defective gene. So, that is exactly the same thing we are doing.

We have used markers and you are looking at which allele of a given marker is invariably present in the affected and whether it is invariably absent in the unaffected, assuming 100% penetrance. You can see here this 35 marker is present in all the affected individuals and it is not present in any of the unaffected individual, suggesting this marker could possibly present close to a gene which is defective, because this marker itself is not causing the disease, because these are repeats which are not part of any genome, any gene. So, they are present randomly and they may not have any functional significance. That is why you see a large number of variation in the population, but they may be present close to a gene that is what shown here.

(Refer Slide Time: 41:44)



For example, this is a marker that let us say that has got number 35, right? The allele that causes the disease and close to this marker in the chromosome very near by you have a gene that causes a disease which you call as HD here and because they are present close to each other, now they are going to go together. They will be not be separated by a recombination; they will not be separated by a recombination. Therefore, whenever you see the individual is affected, you will have that marker, like you can see affected and the marker is present. In those individual who are not affected, this particular allele of a given marker is not present. So, that is what you do. You basically look at a given marker, an allele of it that is co-segregating that is present in all the affected and not present in the unaffected, if the penetrance that you are talking about is 100%.

So, this is the way you go about and identify a small chromosomal region that could possibly be contributing to the disease.

So, how do you really calculate? Because you are going to look at thousands of markers to begin with, because you are looking at all the chromosome, when we talk about autosome, you are having 22 chromosomes, each chromosome you may have 100 markers you are going to look into a large number of markers. So, you need, you know, here we are doing it with a some statistics and then try to sort of guess whether this is the region that could possibly harbour the gene. So, that is called as a likelihood, the possibility.

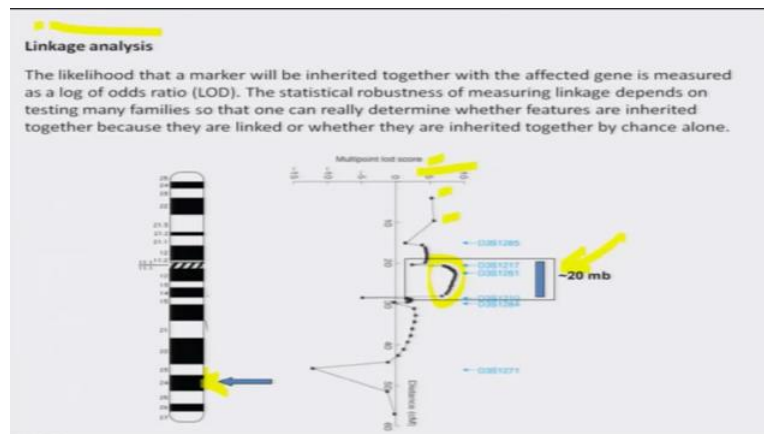
(Refer Slide Time: 43:20)



So, the likelihood that a marker will be inherited together with the affected gene is measured as a log of odds ratio. So, this is a statistical approach people use that you find a marker co-segregating with a phenotype, because it is present in close proximity to a gene that is defective and is not happening by chance. It is not random, you know, probability, right? It happens likely, because it is

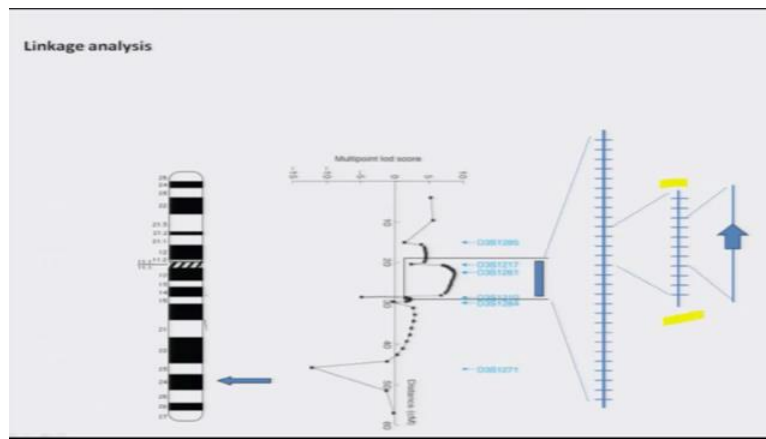
linked, closely linked to that particular affected gene. So, that is what you calculate by using a statistical method and that is what depicted as LOD score, LOD that normally we call. So, if the LOD score is higher, then it is likely that the gene is present very close to that marker. If the LOD score is very low for a particular marker in your analysis, then it is not likely that the gene is present close to that region.

(Refer Slide Time: 44:19)



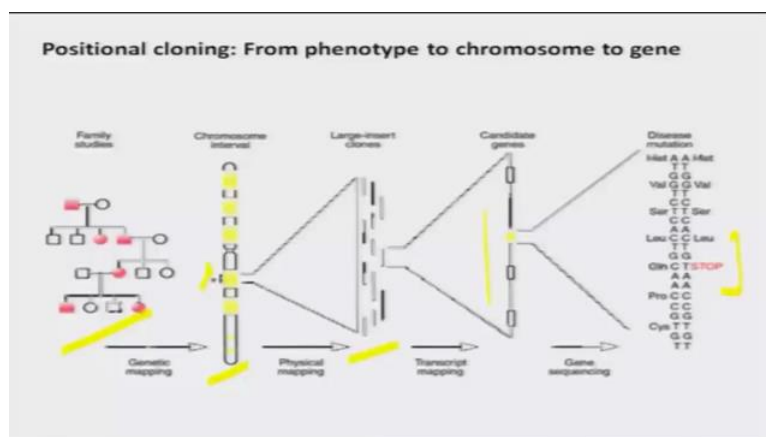
So, this is how you come up with, for example what is shown here is, there are, each square here is a marker that was tested for the so called linkage analysis. That is what shown here that you have done a pedigree and done the genotyping for all the individuals in that pedigree and then you are calculating the LOD score and you find that a small region of the chromosome you have all the marker showing very high LOD score, like that what you see here. That suggest likely that the gene is present somewhere in this particular region, say for example it is present here. So, from all the 22 chromosomes we have come to a small region of a particular chromosome by looking at markers. They themselves are not genes, but because they are present somewhere close to a gene that is defective they would co-segregate with a particle phenotype. So, that is how you are able to narrow down the gene. So, that's how you narrow down a region, which possibly could have the gene. So let us see how from that segment, that small segments that you identified by this linkage analysis, how do you now go and find the gene? So, what is the second step?

(Refer Slide Time: 45:30)



The second step is that the region that you identified narrowed down what you call here you need to look at large number of markers in that. So, that is what the next step is. Now, you do not really worry about the rest of the chromosomes or even the rest of the regions of this particular chromosome. You know this segment you go and look at large number of markers. If markers not available you generate markers and then look at again, analyze, you know, you do this, you know, genotyping and then you narrow down further. We will talk about how do you narrow down and then finally once you have done that this is the smallest region that I can define, then, you go about looking at what are the genes that are present there. Say, suppose you found only one gene, then go about checking for mutations in the gene and if that is having a mutation, then you have found the gene. So, this is the step that you go about doing it.

(Refer Slide Time: 46:26)

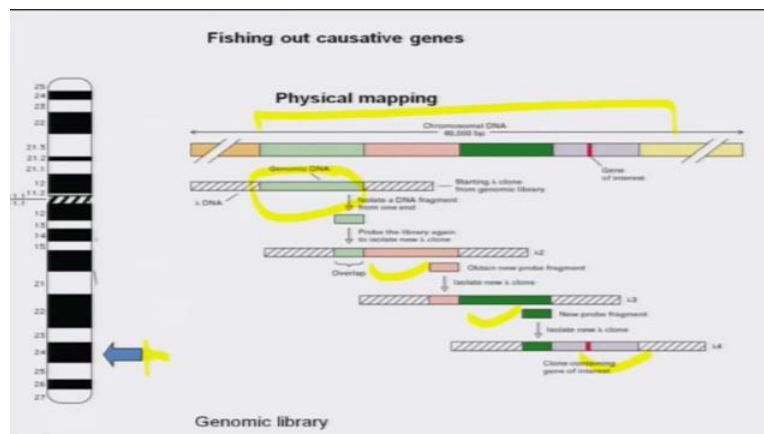


That is what shown here in this cartoon as well. So, you start with just a recap, because you are getting into a problem that you, I would help you to solve and later you may have some assignments given on that. But, what we have done is that we have used a pedigree approach to identify whether it is a autosomal or sex linked or X or Y or dominant, recessive. Then, you have done the genetic mapping, you have used large number of markers and then you are able to say this region of the

chromosome, possibly the gene is located and then you go about characterising that region of the chromosome to find the genes, right? What is shown here and then look at whether a given gene is having a mutation something like what you have seen. So, in a patient if there is a mutation resulting in stop codon, then you know that you have gotten the gene that you are looking for.

So, in this process what you have done is you really did not bother about what is the gene, what is the function, whether it is expressed, what kind of protein it codes, you did not bother about that. You looked at what is a chromosome that is linked to the disease, what are the genes that are present and then, whether any of the gene carries any mutation that could be causing the disease. So this is an approach based on the position of the gene. Therefore, it is called as positional cloning, because more often this led to a discovery of a new gene, more often that is the case and therefore, it is called as cloning, because you identified a new gene. So, that is why it is called as positional cloning. So, let us see, you know, in this, in this particular slide what I am saying is that if you have identified and narrow down a small segment of chromosome you need to look at the genomic region. So, now you have to carefully look at that genomic region, what are the genes that are present there, right? For that you need to go to the library, what you call as genomic library and get fragments that represent that region of the chromosome, arrange them and then, you know, characterize that region as to what are the genes that are present.

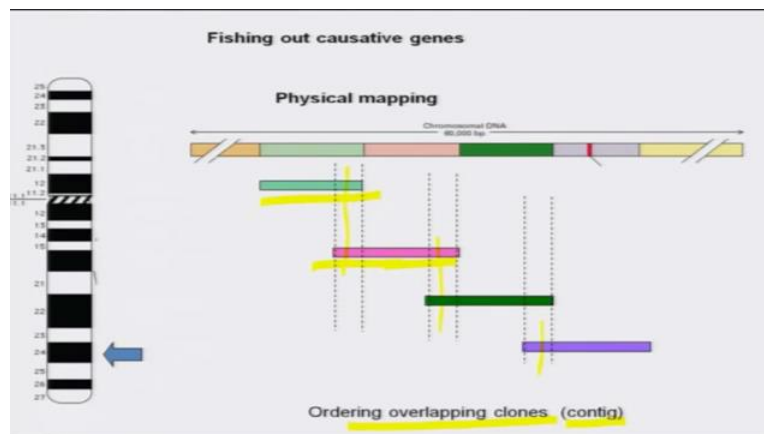
(Refer Slide Time: 48:30)



How do you really do this? This process is called as physical mapping. What do you do in this case is there for example, this is the region you narrowed down, you have markers and then you go and use this marker to screen the genomic library and then once you screen the genomic library you would get a DNA fragment from a library that represent somewhere here, you know, a fragment of that particular chromosome.

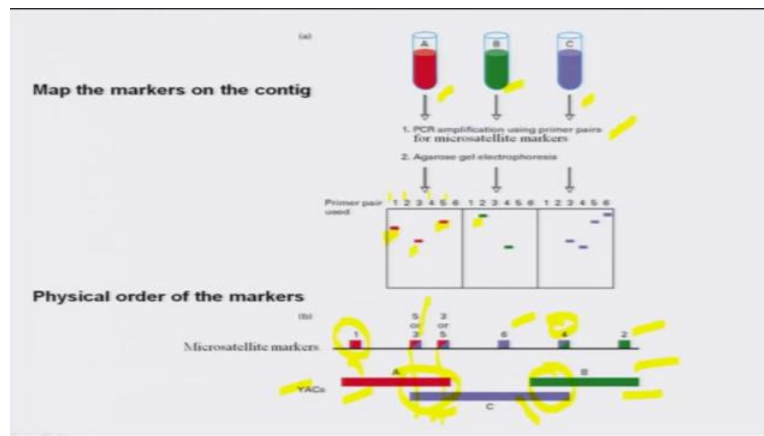
Now you use this genomic DNA fragment that you got from the library to get the clones, different fragments that overlap with each other and representing the region which could possibly harbour the gene. So you have to, you have to take pieces of the DNA that are there in your genomic library, stitch them together to get the DNA representing that region of the chromosome. So, this is called as physical mapping. Sometimes it is called as chromosome walking, because it is a very slow process; you get one, one step, get the other clone that is the second step and then the third clone that is third step. So, you slowly move on either side to get all the clones that represent that region.

(Refer Slide Time: 49:41)



The question is why should you do this? Because, when you do this kind of exercise, you would have a fragment representing a part of the chromosome and that results in the identification of another piece, both of which may share some region. That is the way you are able to say, this follows this. So, this overlap is very very critical. If you can recall our discussion on how do you create genomic library, we said that we do either a partial digestion wherein all the sides are not cut or we use two different enzymes to create fragments that overlap with each other. That is the way you will be able to connect one with the other. That is how they are done earlier. So in each, you know, junction, you can find there are small segments that overlap with each other and you are able to create what is called as overlapping clones, which in colour in a, in a term in genomics called as contig.

(Refer Slide Time: 50:34)



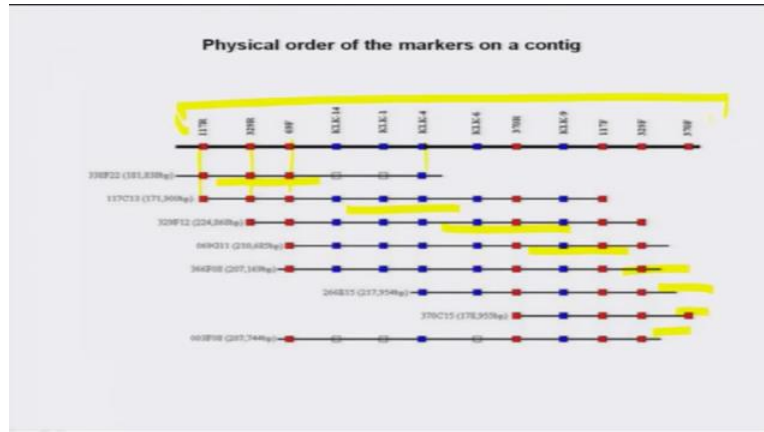
So, how do you really use this contig? So, what is the purpose of the contig? One of the reasons is that the microsatellite markers people used them to screen for chromosomal regions. We know that these markers are coming from this region of the chromosome, but the exact physical order, order of these markers were not known. The reason being the approach that they used for localising a given marker is very different, because then you do not, you know, that time even the fragments were not identified like the way now we are talking about that there is a physical, you know, mapping or chromosome walking. Only in this process you will be able to arrange the markers in the physical order. We know that the markers are present here, but exact order of the markers are not very, very clear and they are very, very essential for mapping studies, like we are going to see now.

How do you really arrange the markers? So, basically what you do is, so you have markers which are nothing, but, you know, the region of the chromosome that you can amplify by PCR. So, once you have got, for example, clones, genomic clones either from yeast artificial chromosome or bacterial artificial chromosome or of phage library, these are the fragments. Now, what you do is, you, each DNA fragment you isolate and then do a PCR and test whether a given marker, you know, is located on this piece. If it amplifies that segment that means it is present there. If it does not amplify that marker is not present. So, as you can see here, you have, different primer pairs, like for example 1, 2, 3, 4, 5, 6 and so on and what you have done for each primer pair, you have used the DNA either from the clone A, B or C and then, you have seen that for some there is a product, for some there is no product.

Based on this information, you are able to place it for example PCR primer 1, you know, that amplified only from this particular clone that is clone A. That means it is located only in clone A, not in any other. On the other hand, some of them are, some of the fragments, some of the PCR products were, you are able to amplify from more than one. So, this information helps you to say the 5 and 3 probably are present in a region that overlap between A and C. Likewise, primer 4, you

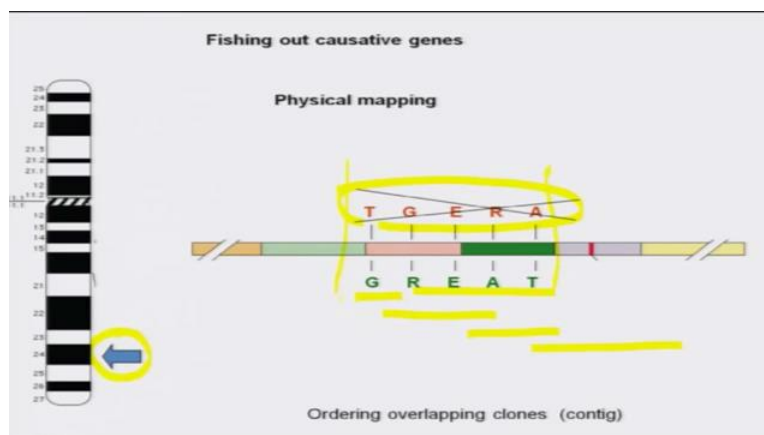
know, is located in a region that overlaps between C and B, whereas 2 is unique to B, 6 is unique to B and likewise 1 is unique to B. So, this is how you are able to position in their relative order. Still you are unable to solve this 5 and 3, which is the order, but if you have had some other clone, probably you will be able to still place them.

(Refer Slide Time: 53:28)



So, this is how you position them and once you are done, you are going to look at something like this, which is called as a contig map. So, all that are shown on the top are markers and each of this line that you see here are the different clones of, for example bacterial artificial chromosome. You can see that these markers are present and these markers are present and so on and you are able to position this is the order of the marker, right? It starts from 11, ends with 30 and so on. So, this is what called as contig.

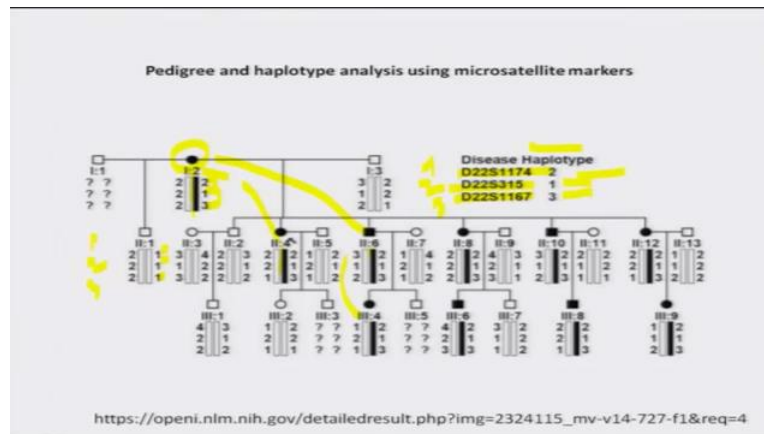
(Refer Slide Time: 53:58)



So to just to summarize what we have discussed, you have looked at a region that you have mapped on to your chromosome likely to have a gene and you want to identify the gene. How do you go about doing it? So, you have done what is called as genomic library screening, identified several

overlapping contigs, clones and you have tested the markers that you have used for the physical order. So, what you have done? You have done PCR using each one of the clones that you have got and you are able to, using this information you are able to say, this is the order of all the markers. This is the correct order. So, you know, all the markers are located within this segment, but you are not sure about the order. Now, you are able to get the order, right? This is the physical order. So, how this particular physical order helps you to identify the gene?

(Refer Slide Time: 54:56)

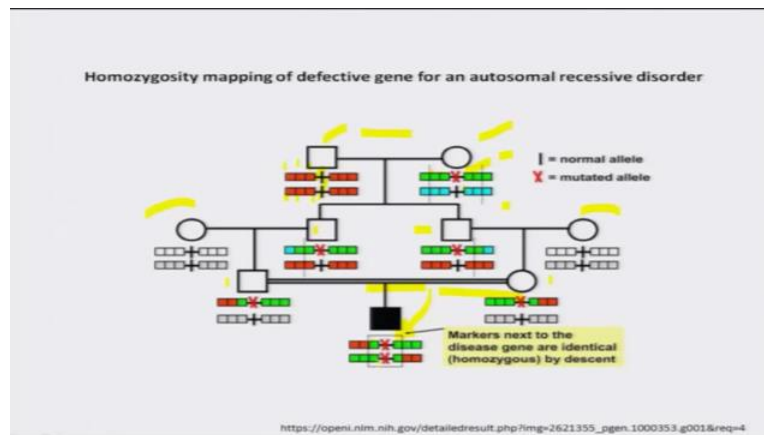


This is a real life situation. This, you know, pedigree that is from one of the papers, publications describing an autosomal dominant disorder, right? You can see here there are three markers. You can see that D22S117, these three are the markers and whatever number that you see here are the allele numbers. For example, this is number 1, this is 2 and this is 3 and these are the three markers. In every individual it is then typed and what you can see here is that there is a particular order of the markers. For example, you have an individual here who is affected and this individual has got 2 1 3, you know, 2 allele for the first marker, 1 allele for the second marker, 3 allele for the third marker and this combination 2 1 3, because these are in the physical order that is also represented by, you know, by the black bar there, is invariably present in every individual that are affected, right?

So, that is, you know, what you are able to get. By now, you are able to clearly look at a combination of alleles that are located to each other what you call as haplotype, you know, the physical order of certain alleles as a block not separated by a recombination is called as haplotype and this haplotype, you can see that goes from this generation to this generation and likewise, you know, from this generation to here and from here to here and so on. So, this you are able to track. Suppose your order was not accurate, you do not know the order, you may miss out this particular information. There were marker somewhere in between these two, you would assume this is not the haplotype. So, for you to identify a haplotype, the physical order is extremely important. Therefore,

it is not simply the information as to where is this allele or marker is located, but the physical order of the markers are very, very important in successfully identifying the regions that could have the disease, right?

(Refer Slide Time: 57:03)

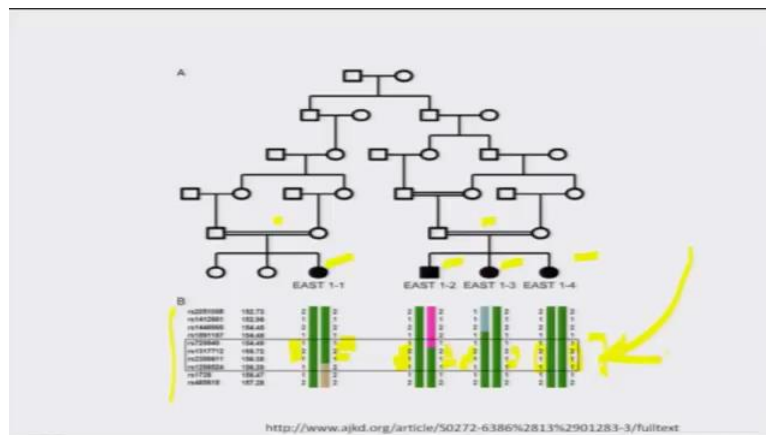


Now, let us see how the haplotype helps us to identify the gene. This is a schematic again to show how do you use these markers to identify the disease gene. Now, you are using a autosomal recessive condition. You have a male, female and we have shown, each box here represent the different alleles, right, of, you know, given marker and here is a gene. This is the gene that you are seeing in between and here is an individual who is heterozygous having a muted allele and she marries a normal individual and she is a carrier and then, they have got two sons, they marry, you know, of course and then in the next generation, his, you know, this particular parent's grand son and grand daughter they marry together, right? It is common, happens in many societies and as a result, this individual who was a carrier for the disease linked allele mutation, now she is able to give this allele to both her son and daughter who are heterozygous, asymptomatic normal.

But now, when there is a marriage between these two individuals, what you see is that, you know, these two alleles could come together in the next generation. When it happens, you become, you know, a homozygous for the mutation, therefore you develop symptoms. Now, you can identify this region, because it remains homozygous, by looking at the markers that are flanking here. You see that these are like the puzzle piece. We have given a different colour. The allele green is one allele, the blue is another allele, red is another allele and you can see that in every generation there, at times there are recombination but you can see around the gene you invariably have the green allele, meaning they are not separated by, you know, the recombination.

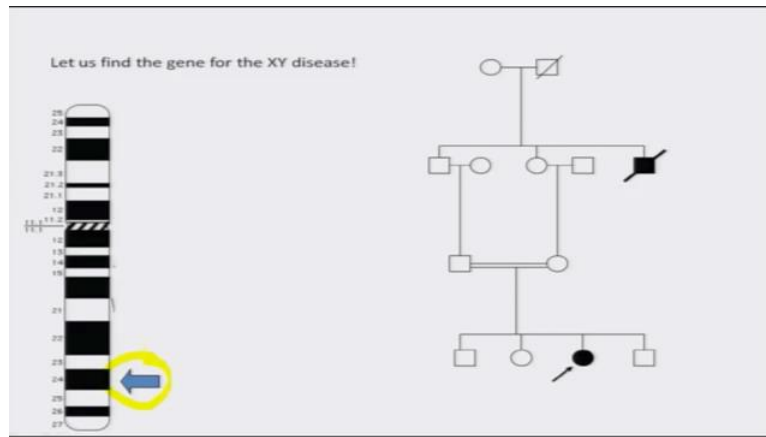
So, if you are able to look at the allele, which allele, you know, segregates in this phenotype, in this family and whether there is a homozygosity for a given allele, invariably every affected is homozygous for that particular allele. That would give you a clue that a gene could be located here and that is what called as homozygous mapping, when you look at a recessive disorders, because it is normally as is shown here, it is a same defective allele, when it comes in the homozygous condition you end up having the disease, because most, most often these are consanguineous marriages, marriages within related and that is a result you see here and that is again shown here with a real life situation for one of the disorders.

(Refer Slide Time: 59:47)



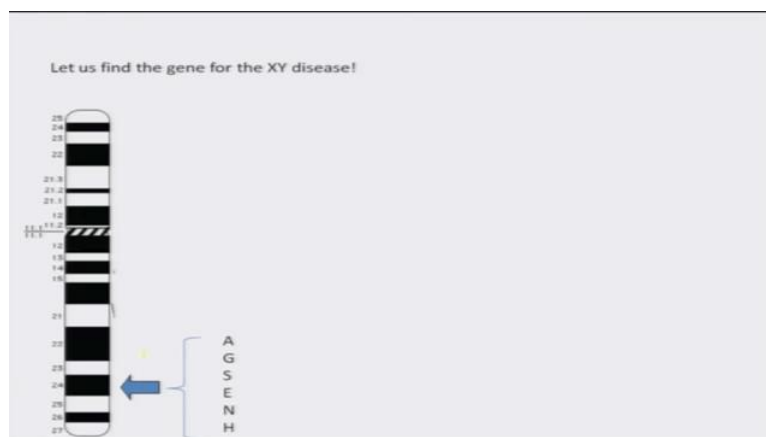
Let us not worry about what disease it is. These are the markers, right and then what you are showing is that there is a homozygous region. This is a large family. You can see there are four affected and there are consanguinity, recessive disease and you see a region that is shown with a box and you will find homozygosity, right? These are affected and this 1 1, 2 2, 1 1, 2 2, 1 1. Again, see homozygosity 2 2, 1 1. So, you can see that 211 is present invariably in homozygous condition. That means your gene should be located somewhere within this segment. So, this is how you arrive at the gene. So, that is what you call as homozygous mapping.

(Refer Slide Time: 1:00:32)



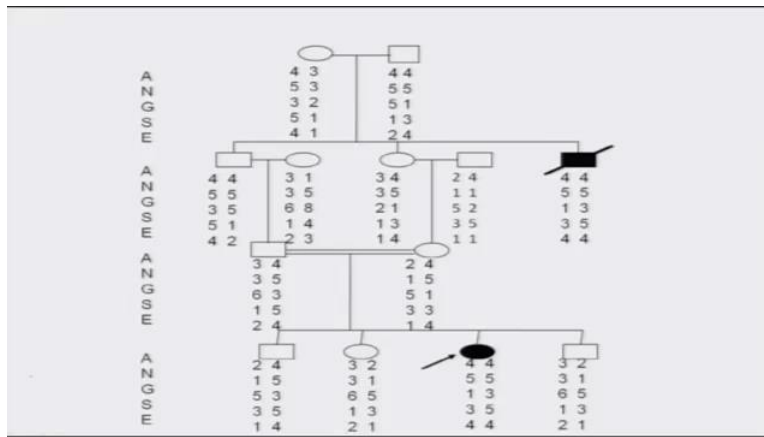
So we are going to look at one such problem. So, I help you with, you know, solving this puzzle and then we may give of couple of examples for you to work out on your own. So, let us find a disease, we name it as XY disease, nothing to do with X and Y chromosome, some name and this is the family. It is clear from here that is a recessive disease, because you have an affected, you have consanguinity, affected parents are not affected and of course, there are relatives who are affected here, right and we have already done the whole genome, large number of markers and we are able to narrow down a region of a chromosome which has got a gene for this particular disease, right? What we need to do? We need to narrow down the region, we need to identify more markers, you know, make them in physical order and look at the homozygosity, as to, because this is a recessive disease. That is what you are going to do. So, these are the markers, right?

(Refer Slide Time: 1:01:34)



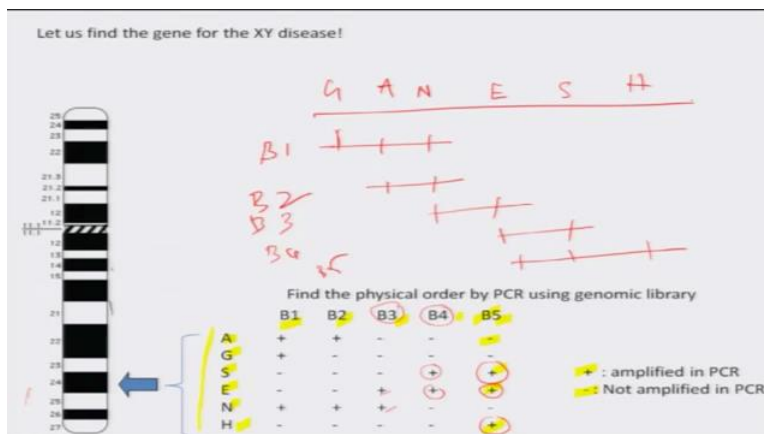
The markers are A G S E N H, right? Some names that that you need not really worry about, right now and these are the markers and what you do?

(Refer Slide Time: 1:01:44)



You have done genotyping. For every marker, you have done a PCR and you are able to identify the alleles and that is what shown here. You can see that quickly. There is no large extended homozygosity or any thing. This is a kind of thing that you see that that you are a, you are getting alleles from both the generations, going to the next generation and so on.

(Refer Slide Time: 1:02:08)

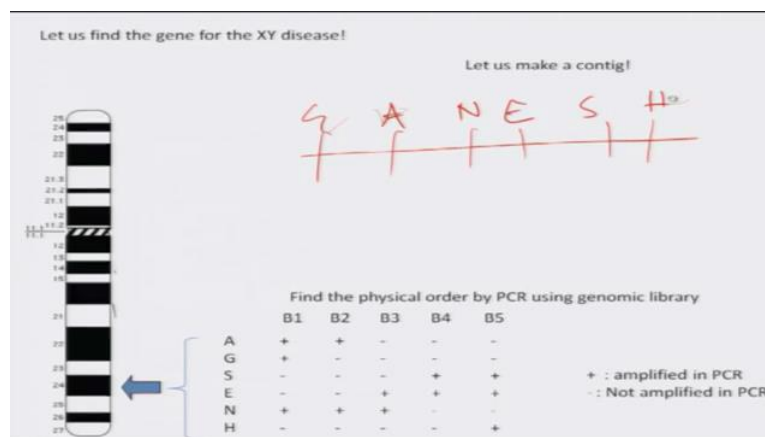


So, what you need to do is that you need to identify the correct physical order of these markers. So, what do you do? You have to go for the genomic library and then identify clones that represent this region of the chromosome and do PCR for each of the clones and then identify the order, right? So, what you have done? That is what you have done. So, you have done that, you have, so these are the markers A G S E N H and you have done a screening in the genomic library. You are able to get five clones, say these are bacterial artificial chromosomes; so, therefore you call as a BACS B1, B2, B3, B4, B5. Here, when you say "minus" that means the PCR did not work, meaning that marker is not present in this clone, whereas S marker and E marker and H marker are present in B5. So, that is the indication, so plus amplified in PCR did not amplify, right? So, this is what you use, right? Now, this is the information that you have.

Now, based on this you have to arrange the order of this particular set of markers. How do you do it? So, let us say I am going to put is that, I am going to say this is B1 clone. The B1 clone is positive for, as you can see here that it is positive for marker A, G and N. B2 is positive for A and N. Therefore, I would assume it is, G is away from A, it is not in between A and N, therefore it is away from it. So, let me say that A and N are close to each other. So, I would, you know, put a line and you would say that here it is, this is A and this is N.

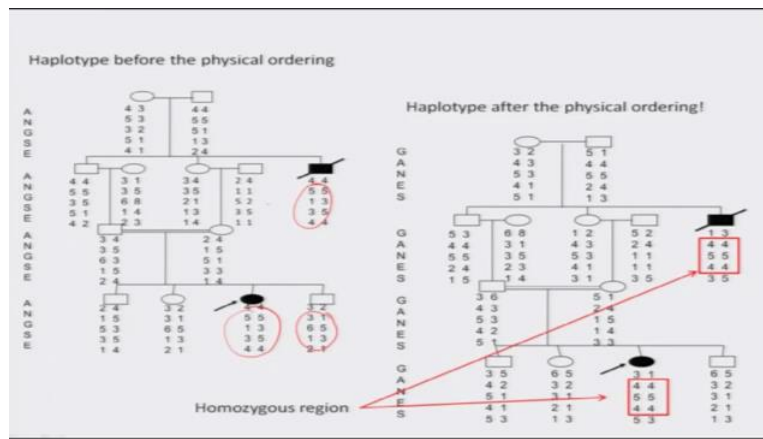
Let us look into B2. B2 is again positive for A and N and I would say this is B2. Now, let us look into A again. So, sorry, let us, let us look into the B1. B1 is also positive for G and no other marker has got that G marker, right? So, then I would put it on the extreme left, because they are not present in any other. So, I would say this is present here. Let us go to B3. B3 is positive for E and N. So, N is already there. I would say E is here and then B4 that is positive for S and E. So, E is here, therefore, S should be here and that is like this. This is B, this is B3, this is B4 and then you have B5, which has got S, E and H. So, what is missing? H; it has got everything. So, when you order, you are able to get an order which also convey some meaning. For example, it is my name. So, it is jumbled up when you order, it, it gives you the, so that sort of tells you what is the physical order of these markers. Now, why should I do this? How does it really help me? That is what you are going to see.

(Refer Slide Time: 1:05:34)



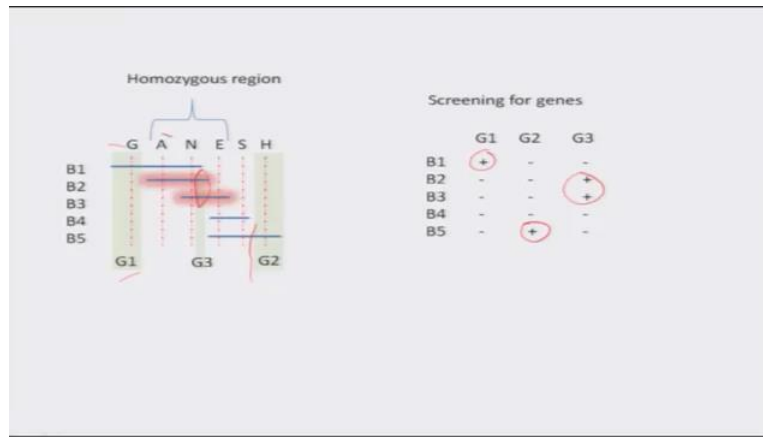
So, you know, by, by doing all these things I am able to make a contig, which clearly tells me that this is the order of the marker that is GANESH, right? That is the contig, right, we made it.

(Refer Slide Time: 1:05:51)



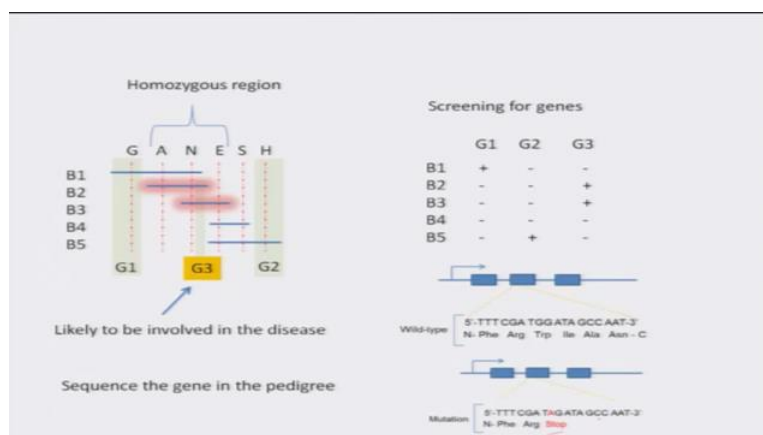
Now, how do they really help me? So, this is what the genotyping that I had. For every marker, I have done the amplification. This is what it is. Now, this was before, the haplotype was before the physical ordering was done. Now, we know that this is not the correct order. We have changed it, right? So, now what I have done? I have rearranged everything based on this particular marker, right?

Now, what you need to look at? Because it is a recessive disease, you have to carefully look at the homozygous region. So, if you look into these two individual that are affected and you will find markers A, N and E, you know, they remain in homozygous condition. You see that 4 4, 5 5, 4 4, 4 4, 5 5, 4 4 and you do not find that homozygosity in any of the affected individuals, sorry unaffected individuals. So that, such is that this is a region likely, you know, having the disease causing the, you know, disorder. This is the region that is likely to have the gene that causes the disorder and if you look in here, you know, when you have not done the homozygosity mapping, when you have not done the correct ordering of the markers, they are not really explaining what is the real scenario, which region is showing the homozygous. That is why the physical ordering is very, very important.



Now, how does it really help you? This is the contig region, homozygous region that you have seen that is A N E, here. This is a region that remains homozygous. How does it really help you in identifying the gene? The way it helps us is by screening for genes. Now, you need to look into what are the genes that are located here, right? So, I know there are three genes G1, G2 and G3. These are the three genes. What did I do? I did screening for each of the gene, whether it is, this gene is located on B1 or B2 or B3, B4, B5 and likewise, it is amplified here, amplified here, this particular gene amplified in two different BACS. Now, you had used this information to locate again the gene BAC in the contig and that is what you see here. G1 amplified only in B1. That means it is present in a region that are not shared by any other BACS. Likewise, G2 is located only in B5. Therefore, their position is a region that is not shared by any other BAC clone, whereas G3 amplified by both B2 and B3, but not by any other clone and this is the BAC and you know, this is the only region, you know, which is not present in other BACS.

(Refer Slide Time: 1:08:43)



Therefore, that is where the G3 is and likely, because that is also the region where you have homozygosity and therefore, you will be able to confidently tell and this is likely the gene that could cause the disease and is involved in a disease that you are looking at and now you go on sequence.

If you are lucky, you can find mutations in the affected and that is how you identify the disease gene.

So, that, you know, pretty much ends the first lecture of week IV and wherein we have looked at the approach people use to identify the disease gene in monogenic disorders. We have looked at how markers help us, how you use markers to identify the haplotype, how do you narrow down, for example in case of recessive disease for a homozygous region and then locate a gene and then identify a novel gene which may cause a particular disease. So, this is the approach people take to identify the disease gene and with that we will end the first lecture and in the next lectures we are going to look into the other aspects like, how the other approach, what are the other approaches because of the post genomic era, meaning you are, you now know the human genome sequence, how that has changed the approach. It has become much more easier now. So, that is something that we are going to look into, in the next lecture.