Introduction to Developmental Biology

Prof. Subramaniam K

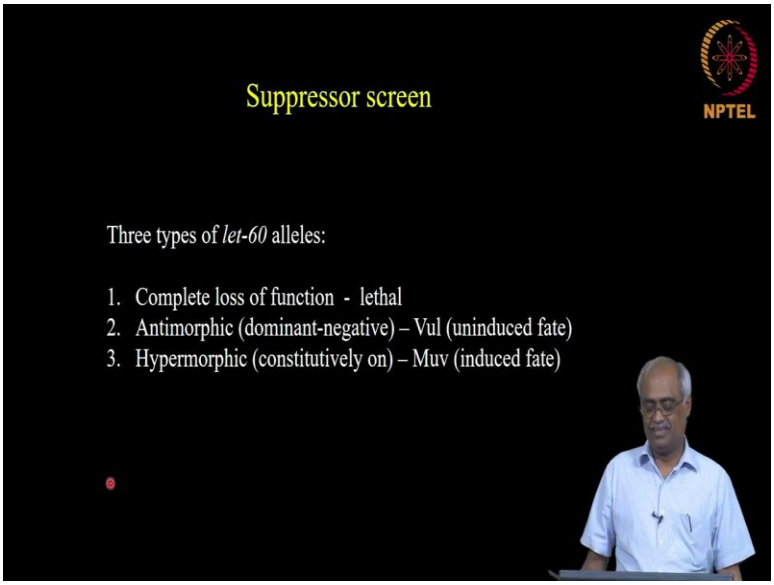
Department of Biotechnology

Indian Institute of Technology - Madras

Lecture-12

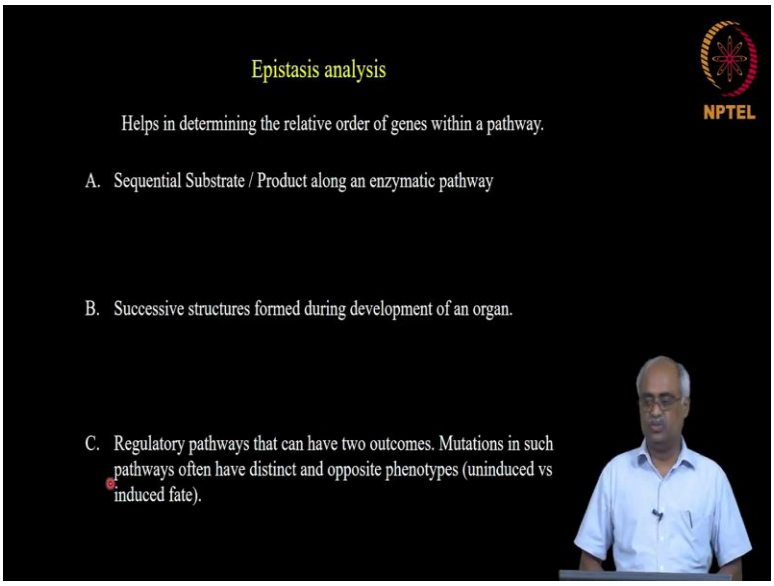
Genetic basis (Part 5 of 5)

(Refer Slide Time: 00:14)



So, in the last class, we discussed modified screens like the suppressor screen and synthetic screen. So, I said that we would not go into other varieties, which are the most commonly used ones. Now we will see some of the genetic techniques that could be useful, but primarily we are going to focus only on epistasis.

(Refer Slide Time: 00:41)



So, epistasis helps us find the order of a gene function if multiple genes are involved in a pathway. One of the most common, well-known pathways would be the enzymatic pathway, for example, the series of biochemical pathways like glycolysis where glucose is converted to glucose-6-phosphate and so on to pyruvate. Glucose-6-phosphate will be formed only if there is glucose. Fructose 1,6 phosphate is possible only if fructose-6-phosphate is present and so on. So series of enzymes act in that pathway.

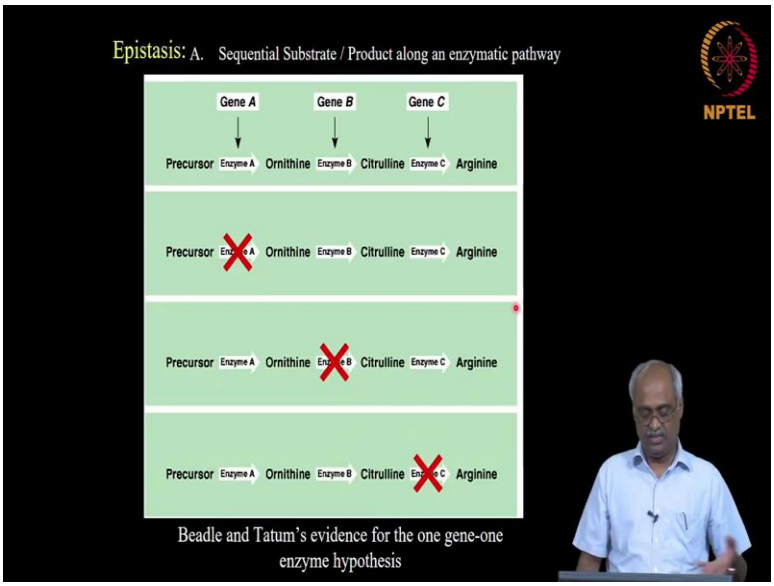
Similarly, the successive structures formed during the development of an organ follow a pathway. Like for example, during vulval development, only if Pn.p's are present, you can have the induced and the uninduced fates; and only if those fates are acquired, different cell types can make up the vulva. So, you have a starting material, and you have intermediate steps, and then you build upon something over multiple steps. So when a limb bud is formed, it protrudes some more, then it spreads into making different bones, and then the web dissolves to make the final structure.

So, these two are very similar, a substrate-product along an enzymatic pathway or successive structures in making an organ.

The third one, the second distinct one, is a regulatory pathway where activation or suppression happens through sequential signaling. So, let us see how to identify genes that function in these

kinds of pathways. There could be multiple genes involved in it; you have intermediate steps; for every step, there may be a protein; therefore, there may be a gene. And the mutations in any one of them like loss-of-function may have the outcome being the same, but you want to know in what order do these genes act, and that is where epistasis helps.

(Refer Slide Time: 04:47)



So let see how epistasis is done. So this is a typical example from Campbell. So we know that Beadle and Tatum discovered the one gene-one polypeptide hypothesis. Originally it was one gene-one enzyme, as illustrated in this cartoon, but eventually, we see that it had become one gene-one polypeptide. So in this cartoon, each enzyme is mutated to see the required precursor and product being made. For example, in this figure, the wild-type grows well if you provide some ammonia, a nitrogen source that acts as a precursor. From that, it can make the intermediates to make the amino acid arginine.

But let us say you have three steps here, and enzyme A is mutated. Then that organism would grow only if you provide ornithine or citrulline or arginine, and providing precursor alone will not be enough, similarly, for the other two mutations. Now let us look at mutations one and two, like enzyme A not being there or enzyme B not being there. Now instead of looking at what precursor you needed to give to grow the organism, let us look at what product formed for this discussion. So mutation in enzyme A will not make ornithine, while a mutation in enzyme B will make ornithine. Now, if I make a double mutant A and B, what will be the phenotype concerning

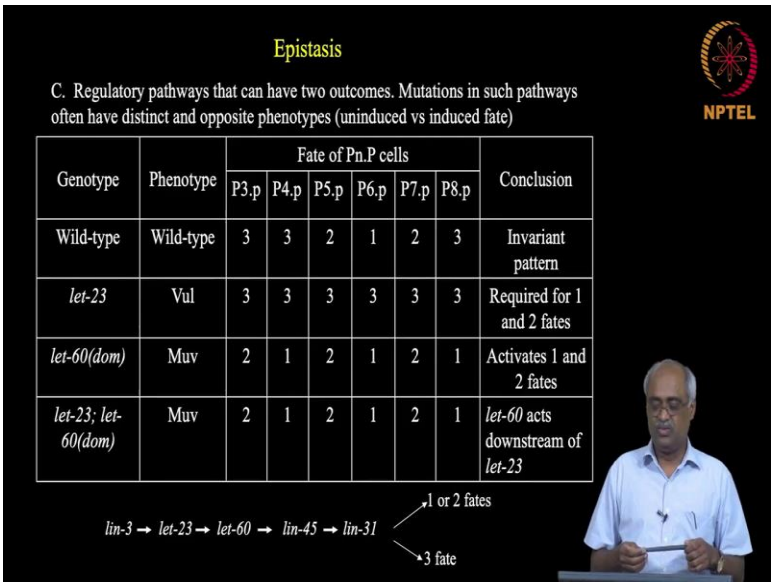
whether it will make ornithine or not? It will not make. So, the AB double mutant phenotype is the same as mutation A; in that context, this is an epistasis analysis. Here you say A is epistatic over B. So whichever gene, whose mutant phenotype prevails in the double mutant, you call that as the epistatic over the other. In this example, A is epistatic over B, so A means no ornithine AB is also no ornithine though B would have made ornithine. So, therefore the AB phenotype is like A phenotype, so you say A is epistatic over B. Here, in this situation, an upstream block is epistatic. So purely based on this phenotype like AB phenotype is like A, and knowing that it is a substrate-product forming pathway, you will say the upstream block is epistatic.

For example, let us say in an automobile manufacturing assembly line, in the initial step, say somebody brings the body parts and put together. Then somebody mounts the engine, and then somebody fits the wheels. Now, if the body part assembly workers have not come means, then the rest of it cannot happen; although the rest of the machinery and workers are all there, you are not going to make a car. So, the first part is essential, and therefore that is epistatic. So, whether the body part assembling workers have come or not, and the wheel assembly workers have arrived or not, the end phenotype is no car is made. But when both are not there again, no car will be made, and if the first people are not there too, it will be the same.

So, the first block will be epistatic over everything else. So, when you see that relationship, then you know that enzyme A must be acting upstream of B, and that is how this picture is drawn. No single mutation will reveal that enzyme A is required to make arginine or ornithine. It will also not tell you whether it is acting upstream of B and the intermediate product converted to the other one. So this is a more straightforward example. If you take a more extended pathway, then you will find the complexity like if you go to wormbook.org and learn about the sex determination pathway there, you will see a series of genes and their opposite phenotypes; therefore, it is good epistasis analysis. So you will find that without epistasis, you cannot figure that out the pathway,

So you will see that in vulva development because we have an example in which we will learn all methods. So this is one example. As I mentioned in the previous slide, A and B are similar, so we saw an example for A.

(Refer Slide Time: 10:30)

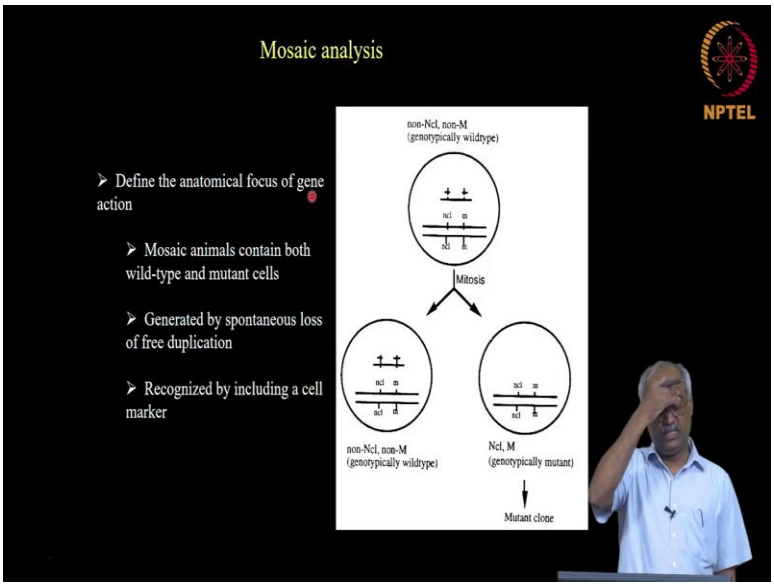


Now we are going to see an example for B, which is vulva development. So by doing mutagenesis and identifying mutants that follow Mendelian inheritance, we learned genes control vulva development to begin. So, now we know that 20 or so genes are involved in vulva development. So how do we know which acts first, and which acts second? So, here is an epistasis analysis. So *lin-26* is vulvaless, and by doing a detailed Nomarski observation, you find there are no Pn.p cells made. So, all you can say is *lin-26* is required to make Pn.p cells. Now you have *let-23* mutant. So, there all Pn.p's became tertiary fate cells, and you are saying it is required for induced fates. Now when you make the double mutant, Pn.p's are not made. So, two things you are learning here, assuming that one did not do the cell lineage. One, here you are learning Pn.p's need to be made, then only you can make the tertiary fate cells. Second and more importantly, for this purpose of this discussion, *lin-26* acts upstream of *let-23*, so if *lin-26* is not there, the previous block happens. So, through this, you find an earlier block is epistatic. So, epistatic means when you generate a double mutant, the phenotype of whichever single mutant is prevailing in the double mutant, you call that single mutant phenotype as epistatic.

In this case, the phenotype of the *lin-26* single mutant is no Pn.p cells and *lin-26; let-23* double mutant is like that of the *lin-26*, and therefore you say *lin-26* is epistatic. And by merely saying *lin-26* is epistatic, we cannot conclude that it is acting upstream, for that you need to know that this pathway belongs to which one of the two categories (A or B). But there is an opposite

situation where I often find people getting confused in the exam, which are the regulatory pathways.

(Refer Slide Time: 13:14)



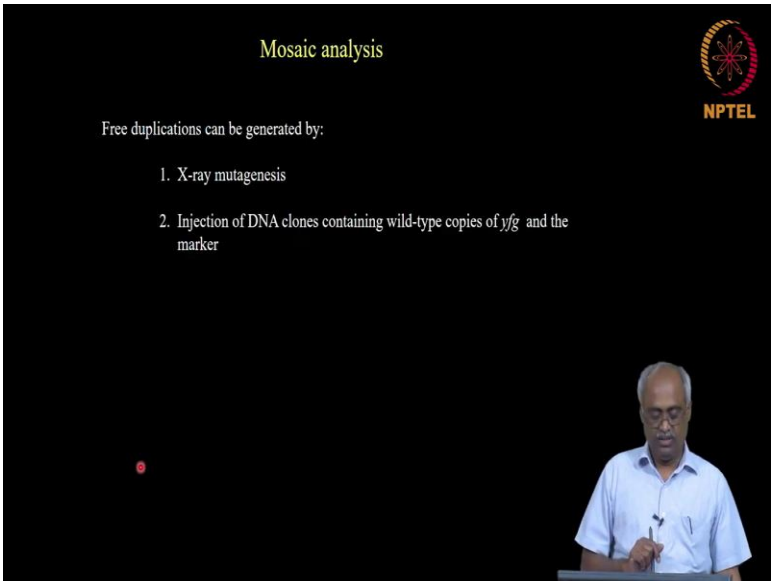
So let us see mosaic analysis, so why it is called mosaic analysis because the organism will be mosaic because some cells in some tissues will have the wild-type function of the gene while others will not have. They would have lost that particular gene. Let us see an example of how you will do this analysis, here you have a mutation **m**, it homozygous for that specific mutation. Now you have an extra copy of DNA usually generated during large scale mutagenesis or using strong mutagen like gamma radiation. So they result in bits and pieces of chromosomes called duplicates, and these duplicates are characterized and are known based on the phenotype they confer. In this particular example, we have a duplicate in which the wild-type copy of this mutation **m** is present; besides, we have a wild-type copy of the marker mutation, **Ncl**. If this cell is homozygous for **Ncl,** then the nucleolus will be bigger in these cells. So now, if the cell loses the duplicate, then that particular cell's nucleolus will be larger because it is not going to have the wild-type copy for **Ncl**. So now we know that the duplicate is lost, which means the gene function also must be lost. Then we can check what happened to that cell or what happens to the descendants of that cell. So this is random because this extrachromosomal DNA is not going to follow Mendelian inheritance.

Since it is going to be random, you do not know in which cell it is going to be lost, and that is why you need this marker mutation. This is one example where **Ncl** is being used, but there are varieties of mechanisms to do this. Like for example, we learned about Cre-lox earlier when we learned about the enhancers, so there Cre activation can be used temporarily or conditionally or tissue-specific, and so on.

So now, after mitosis, some cells will maintain the duplication; thus, that cell will behave as wild-type because it is **non-Ncl**; therefore, we will assume that it is a wild-type. And the cells that do not have that duplicate will have an enlarged nucleolus, which shows that it is genotypically mutant, and then you see what happens to that cell or the descendants from that cell. So, the loss of duplication is what we are using here as a marker.

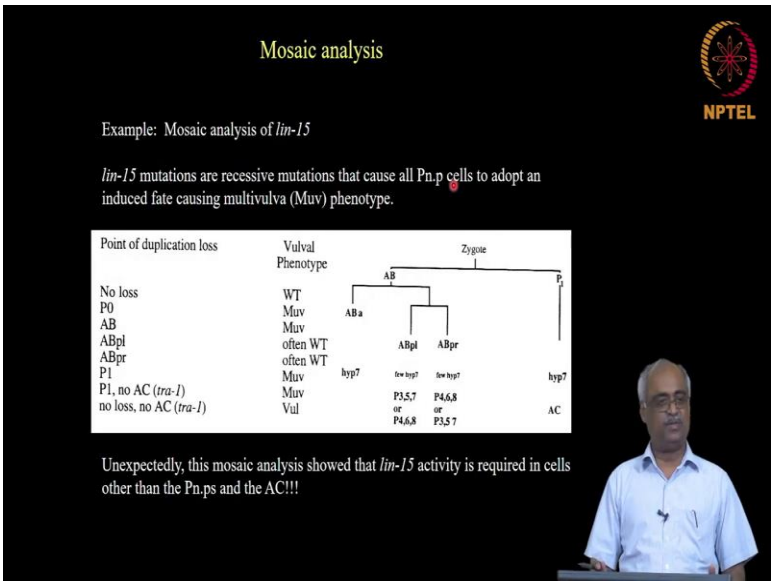
So this duplication is one way of doing mosaic; there are other ways like Cre-lox is another way.

(Refer Slide Time: 16:39)



So, these duplications are generated by X-ray mutagenesis, or you can inject DNA, which can exist in the nucleus as an extrachromosomal copy. So in this slide ***yfg*** stands for your favorite gene. So, the duplicate could have come from an organism like *C. elegans* where duplications are available for different regions, or it may be an injected DNA that has the two markers. So, ***yfg*** and ***yfp*** are commonly used terms to say your favorite gene and your favorite protein.

(Refer Slide Time: 17:23)



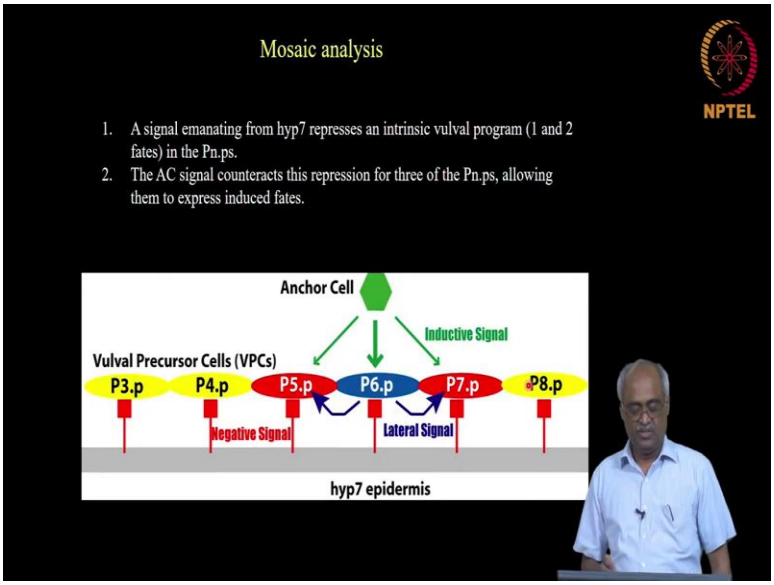
So, here is a real-life example of mosaic analysis. So, here we are going to look at the vulva. So, we learned these genes like *lin-15*, *let-60*, *lin-23*, etc. So, where are they required? At what step these genes are necessary and to identify that we can use mosaic analysis. So, here we are doing mosaic analysis for *lin-15*. So, the *lin-15* mutation is recessive mutations that cause all Pn.p cells to adopt an induced fate that is multi-vulva (Muv). So, now let us look at what happens if we provide a duplication having a *lin-15* wild-type copy and a marker. Let us consider **Ncl** as a marker; in the slide, if you see, we should follow from the zygote and see at what point the duplication is going to be lost and what will be the phenotype of it. So, now you start with the zygote. Let us say in the entire embryonic lineage the duplication is not lost, then you will not find a cell where the nucleolus was enlarged. So, that will end up developing into wild-type, where it will make one single functional vulva.

Now you find an embryo in which the P0 that is in zygote itself it had lost; like AB, P1 both had enlarged nucleolus, then it is going to be *lin-15* homozygous, so multi-vulva. Now we will look at another embryo where zygote was normal, P1 nucleolus was normal, but AB had an enlarged nucleolus and all the descendants from that AB as well. Now that embryo will grow up as an adult with multi-vulva, indicating that its function is required either in AB or one of AB's descendants. Similarly, if you go on to ABpl, the ABp divides into ABpl(left)and ABpr(right), and there if you lose it, often, it grows as wild-type, meaning *lin-15*'s presence in one or the other seems to be enough. Now, if you go to P1, you are losing the Pn.p's. So when you lose in

P1 also the development gets affected; it is not surprising because anchor cell is essential for vulva development, and anchor cell comes from P1, so it is required in both the AB and P1. Now, if we look at another situation wherein P1 is lost, and the anchor cell is absent, but still, you have multi-vulva forming. And then you go to another case where there is no loss and no anchor cell, and there you have no vulva at all, because the anchor cell is the one that induces the Pn.p's to make vulva. Even without an anchor cell, if you are getting multi-vulva, it indicates that *lin-15* seems to be expressed in cells other than the P lineage and anchor cell. So, this kind of finding comes from this mosaic analysis. So, these are the kind of conclusions that comes only from the mosaic analysis; that is why this is taken here as an example.

So this was surprising for people, and then they learned that all these cells unless otherwise, a negative signaling acts to suppress, they will get into an induced fit.

(Refer Slide Time: 22:05)

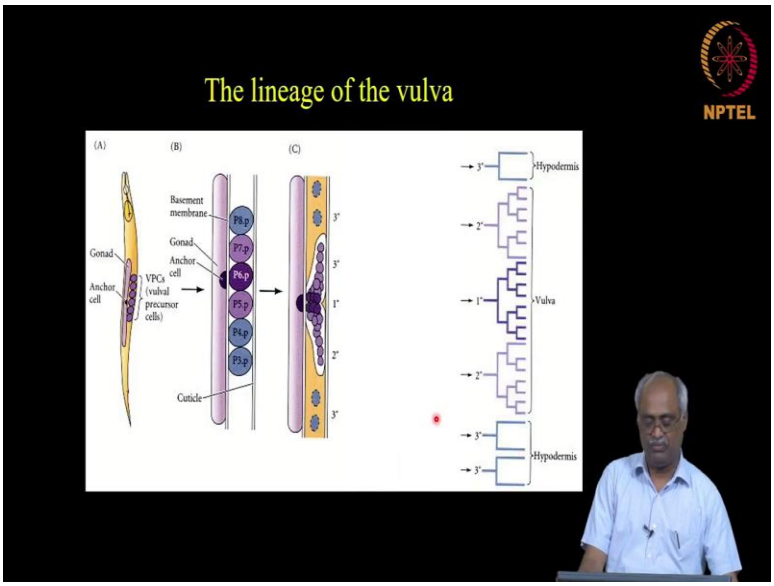


That is cartooned here. So here is the underlying hypodermis, from which there is a negative signaling that is making the Pn.p cells not to get induced at all. So the negative signaling is temporarily relieved by the signaling from the anchor cell, which is stronger on P6.p, and as a result, it acquires primary fate, a slightly weaker signal from anchor cells make the P5.p and P7.p to acquire secondary fate. Therefore only these cells are induced, and the rest are not induced because the negative signaling prevents it. When this negative signaling is lost, here in the *lin-15* mutant, all the Pn.p cells acquire primary and secondary fate. So, all you are doing is just

following the lineage, and then you see where the wild-type copy is lost and what the outcome is. And by knowing this, you can determine in which tissue at what time a gene function is required for a particular structure to form. So, this is the mosaic analysis.

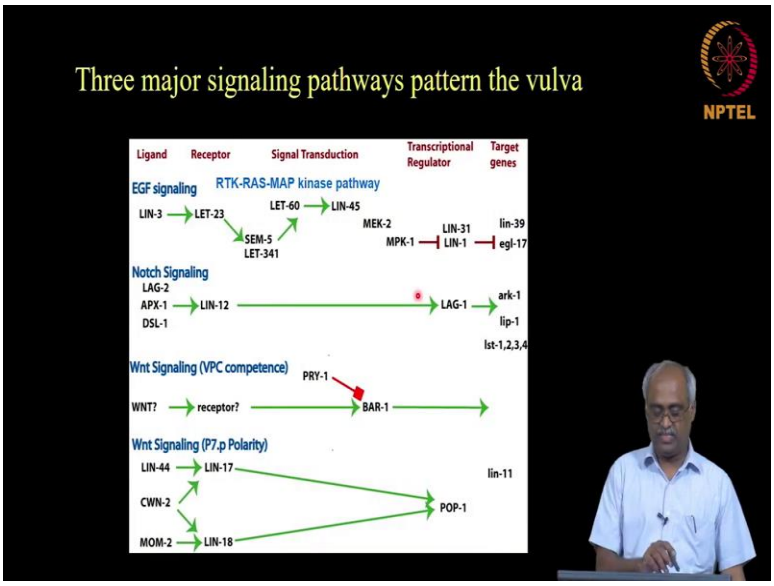
So if it is not clear look at this at a less leisurely pace, then you will get it clearly; all you need is very carefully follow the lineage in the previous slide. So, this completes whatever we wanted to discuss. So we saw the Mendelian laws and then the concept that recombination frequency informs us about the genetic distance and genetic mapping. And how do you jump from that to physical map and identify an ORF, and then what are the genetic tricks you can do in terms of learning more about gene function like epistasis and mosaic analysis.

(Refer Slide Time: 24:06)



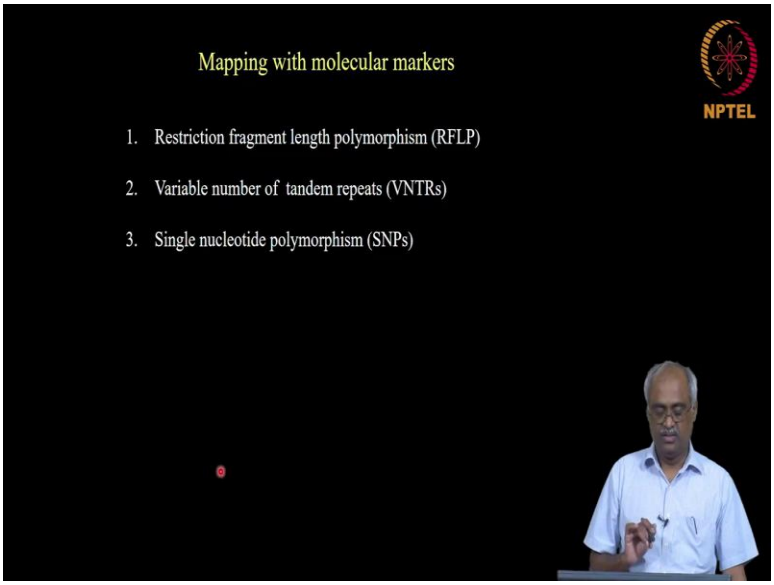
So now, we will move away from this analysis, but before we go, this is the summary. These are the kind of experiments, no biochemistry; no molecular biology led it to this understanding that is shown here. So it was identified that anchor cell signal P6.p through the RAS-MAP kinase pathway, and the lateral signaling in P5.p and P7.p happens through the Notch-Delta pathway. So the first evidence of lateral signaling like membrane-bound ligand, membrane-bound receptor; the idea comes from this analysis. So, much later, we will learn about a model for lateral signaling when we are going to learn cell-cell signaling also. So, this is just a summary.

(Refer Slide Time: 25:01)



And these pathways, like multiple pathways, function to make this one little organ were all identified primarily through the techniques that we have so far learned. So the various pathways involved are EGF signaling, NOTCH signaling (It works in making sure the P5.p and P7.p remain as secondary fate and they do not become primary), Wnt signaling, and another incident of Wnt signaling; all of this comes from these kinds of approaches.

(Refer Slide Time: 25:35)



So next, we are going to learn a few recent techniques to do the genetic mapping. The message is that you do not always need to have known markers with known phenotypes that might limit the ability to do mapping because you may have specific loci where you will have problems with the survival of that organism. So, let us say you have a marker locus and your locus, if you make

double mutant, then they are going to be dead. And the other known markers might be are far away, and you are unable to refine your map. Those issues will happen. All markers are defined by their known phenotype, and some phenotypes may grow slowly or maybe embryonic lethal, so such limitations exist. But if you go to the DNA sequence itself, you are unlikely to have those issues because you are only looking at the wild-type sequence. But then there are subtle sequence variations that can be exploited.

So, primarily we are going to focus on SNPs because this is used most commonly nowadays. Still, historically people used RFLP as well as VNTR (variable number of tandem repeats). In chromosomes, you might have a specific sequence that repeats multiple times; in one strain, you might have 10 repeats; on another one, you might have 20 repeats. So those variations in tandem repeats can be used as a marker.

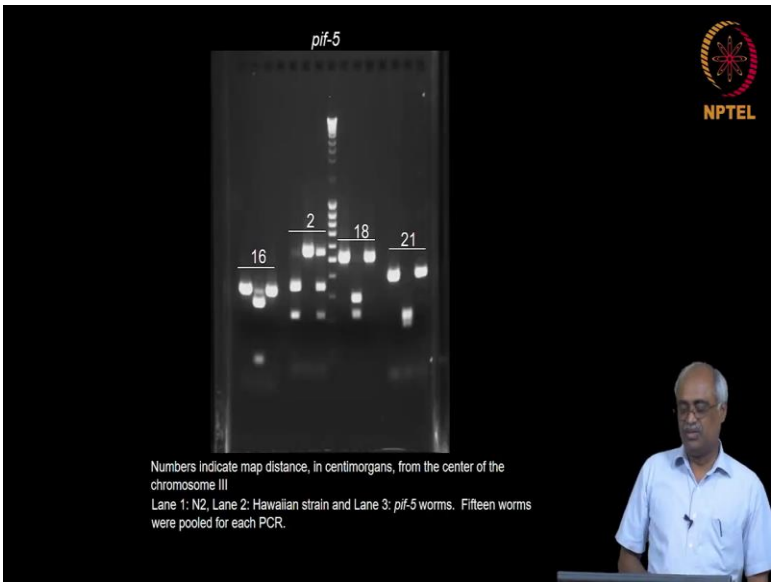
RFLP is a restriction fragment length polymorphism where for example, in one strain, you have a particular restriction site sequence for a specific locus. In another one, a single base is changed, and due to that, the restriction locus is not there. So, between the two strains, you will have varying fragment lengths. Let us say in one strain, there is an EcoRI site in one position, and the next is 6 KB apart. In another strain from that particular EcoRI locus, the next is 3 KB apart. Now when you digest the strain one's DNA, you get a 6 KB band. The second one, when you digest, you might get a 3 KB single band. So, the restriction length is varying, and therefore, there is a change in size. So this is coming from restriction fragment. So that is why it is called a restriction fragment length polymorphism.

Let us look at single nucleotide polymorphism because it is practically beneficial and potent; it is limitless. So, what are SNPs? Genome sequence variations between two different strains that belong to the same species are SNPs. So if look at the genome sequence of two different strains, you find, in a given position there is a change in nucleotide; like let us say the 134^{th} nucleotide on chromosome 1 is A. In another organism, the 134^{th} base on chromosome 1 from the left end is G. So, these two are SNPs. So, the variation is in one nucleotide. So in an organism like *C. elegans,* the genome has been thoroughly studied, and each base is accurately known. So by comparing wild-type strains that exist, you can find SNPs. Like here, the commonly used strain

is the one isolated by Sydney Brenner called N2 that is from Bristol, England. And another strain was isolated in Hawaii. So they are far apart; they never had a genetic exchange at all over millions of years. So these two strains are commonly used. So we would have usually generated our mutation in the N2 strain. If you cross it with a Hawaiian strain, you will get a heterozygous progeny, and now that will be F1. Now you go to F2, let us say in F1 you have **m/+**; so **m** is in N2 genetic background, and the **+** is in Hawaiian genetic background. Now in F2, I will be able to identify 25% worms homozygous for the mutation. Suppose let us assume mutation is on chromosome 1, and in F2, we found one-fourth of the population having the mutant phenotype. If I go to chromosome 2, what is the probability that one of the alleles is from N2, and another allele is from the Hawaiian strain? It will be independent.

Now let us come back to chromosome 1. Let us take an SNP that is 10 bases to the right of my mutant, base just 10 bases, here I am only looking 100 progeny. Remember, in F2 I am not randomly picking worms; I pick only the 25% worms carrying my mutation. Now I am looking at an SNP meaning; the 10th base from my mutant base differs between the N2 and Hawaiian strain. So now in these worms which are homozygous for the mutation, which SNP will be there? N2 SNP or Hawaiian SNP? N2 SNP is going to show linkage, so that is the basis for an SNP based mapping. So an example is here.

(Refer Slide Time: 32:57)



So, here *pif-5* is one of the mutants isolated in our lab, and the student who was mapping it did an SNP mapping. So, he did SNP combined with RFLP, where the SNP creates a restriction

fragment length variation. Say, for example, I have two primers that flank a given base where SNP exists, let us say I get a 500bp product. In between the product, at some point, I have a base where it varies between the N2 and Hawaiian. Now let us say that variation affected a particular restriction site. Now this PCR product after amplifying if I digest with the enzyme the one where that SNP does not affect the restriction site will get cut into two pieces, and the one where it affects will not be cut, so that is what we are doing here. So, here in the gel picture, this is position 16 on a particular chromosome. So, the first lane is pure N2 strain without any crosses. So those worms are taken, and PCR amplified with these two primers for that particular SNP and digested. Since there is no restriction site, it is not digested. Lane 2 is a Hawaiian strain; it is also not crossed with anything; it is a pure breeding Hawaiian strain. So this is PCR amplified and digested since it has a restriction site closer to one end, it gets cut into two. Now, Lane 3 is my mutant after crossing with the Hawaiian, and in F2, I have picked only the worms that show my mutant phenotype. So we isolate the DNA, PCR amplified, and digest with the restriction enzyme. Now I find it is not getting digested; it is looking like N2. What it means is my mutation is closer to this SNP; how closer? It depends on how many worms I took here. If I took large enough that I accommodated a few worms in which recombination is possible, then I might get a mixture as you see for position 2 Lane 3. At marker position 2, the N2 gets cut into two, not the Hawaiian.

So this will vary depending on the marker. In position 2, lane 3 where the mutant is present, we see both versions. So, in lane 3, there is both digested and undigested product, meaning this strain seemed to be a mixture of Hawaiian and N2, indicating that my mutation is quite away from position 2. If you look at the first set, you might say it is closer to 16. Now you look at the result of 18 and 21; then, you realize it is probably far to the right of position 2, perhaps in 16 to 21.

So, we can sequence the entire region of the chromosome, but SNP mapping is practically a lot easier to do. So in *C. elegans,* for the whole of the length of all chromosomes, we have defined primer pairs, and we already know the enzyme used for digestion and the expected products for N2 and the Hawaiian strain. So the *C. elegans* chromosomes are highly rich with these markers,

so the power of mapping using SNP is limitless. So, many students nowadays, right after their chromosome assignment they go for SNP mapping.

Once you come to such a short interval, all you do is you do the whole genome sequencing, which is easy because you isolate the DNA and send it to a facility. They will send you the mutations in the sequence for a specific genetic interval on a particular chromosome. So they are going to omit all those mutations that are not affecting the protein-coding sequence. Even in protein-coding, they will ignore all the synonymous mutations, and they will only identify the ones in which missense or premature stop or deletion, etc. is present. Then you look at the datasheet, and right away, you will know that a particular gene fits with the phenotype, and then you will immediately know what that gene is. Then you can revalidate it by Sanger sequencing of that region to see whether you have a mutation there.

So that is how you can very quickly map, but otherwise, the basic principle is the same as what we learned initially with recombination frequency. The rule followed is the same as what we use between black and vestigial and then in **cn**. Because you are doing the cross and you see which marker shows linkage with your phenotype. So, there it was a phenotype that co-segregates like how often **cn** is seen with black or not. Here, whether the SNP is seen with the mutation or not, only the detection varies.