Introduction to Developmental Biology

Prof. Subramaniam K

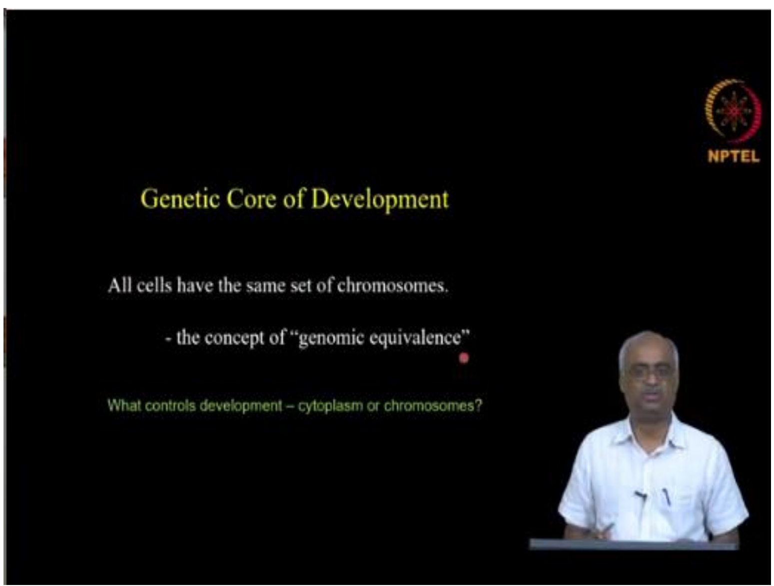
Department of Biotechnology

Indian Institute of Technology, Madras

Lecture No-04

Differential gene expression (Part 1 of 4)

(Refer Slide Time: 00:16)

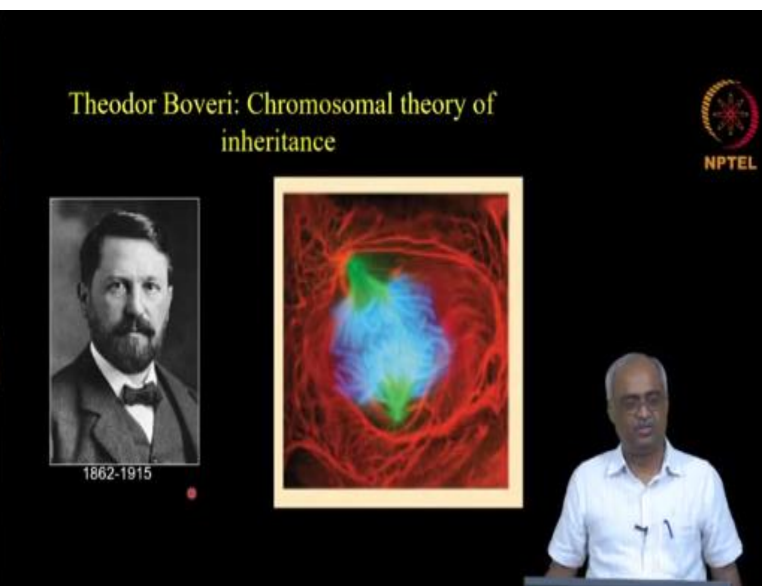


So in yesterday's lecture, the central theme was two things, one how changes in development enable evolution, and second, we looked at cell lineage and fate mapping, the various techniques involved in it. Then I ended up saying that in the next class, let us look at how development is controlled. So that is what we are going to learn today.

So we are going to go through the history, my firm belief is that it helps in shaping our thinking about development. So you might think these are all history, it is over, and what is the point of knowing, but I still feel it is worth taking a few minutes to look at the past to understand how the thought process is shaped up in a given field.

So like epigenesis and preformation, another controversial theory was laid to rest about 20 years ago. So most of you already know all cells have the same set of chromosomes, and that is from where this concept called genomic equivalence comes. What it means is you take any cell of the adult body; they have the same set of chromosomes. There are exceptions like RBCs do not even have a nucleus, and if you were willing to think a little bit more, you must have learned VDJ recombination in Immunology. In response to foreign material, our B cells are capable of generating an antibody that is customized for a particular pathogen, particularly to the shape, not even to the pathogen. So it is being made possible when their genome undergoes rearrangement again and that is called VDJ recombination. So each B cell, therefore, has a genome that is different from another B cell leave alone another somatic cell. So there are variations, but by and large, most of the somatic cells, in any organism has the same set of chromosomes, and that is what we call as genomic equivalence. So now, the question is, what controls development?
Is it the cytoplasm of the oocyte?. Oocyte comes with a huge cytoplasm. It is much bigger than most of the other cells. So, does it come with all the necessary things for an organism's development?. So that was one school of thought another one was chromosomes. It is the chromosomes that determine not the cytoplasm. So, then people pursued both of these to find evidence, and that is how science progresses so we will see those.

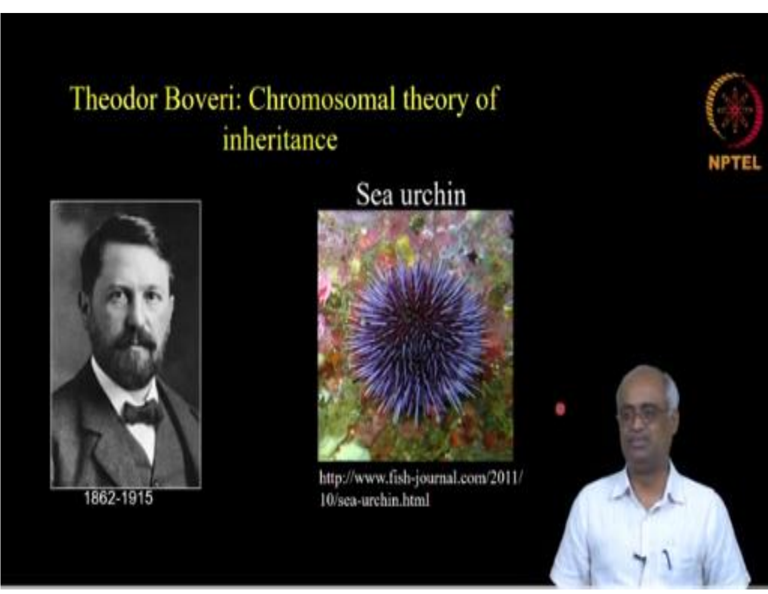
(Refer Slide Time: 03:41)



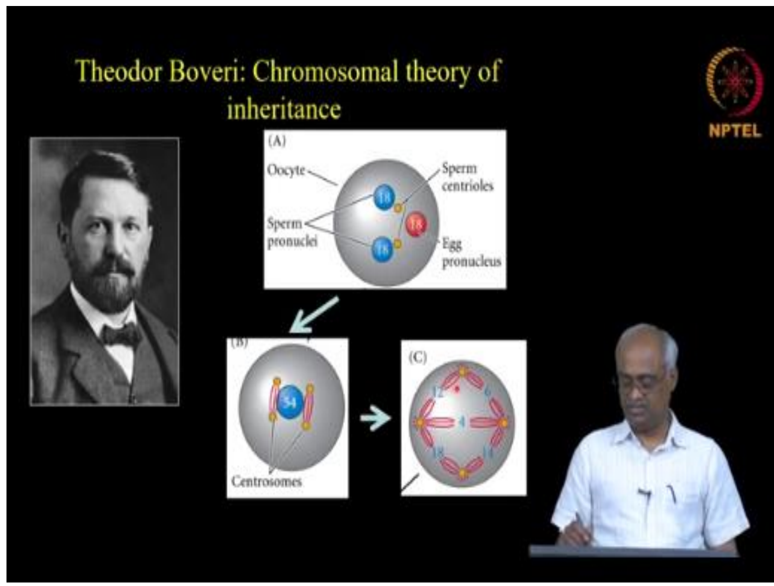
So, one of the famous embryologists who has contributed substantially to the field of development as well as genetics is Theodor Boveri. So, one of his evidence is what we are going to look at that

is he looked at cytoplasm. We are going to skip some of the genetics here that is, how did people find Mendel's genes are located on chromosomes? That we are ignoring. So in genetics class we can learn that, so here we are assuming that we already know genes are on chromosomes. Very briefly one of the things was the behaviour of chromosomes when cytologists looked and observed the chromosomes where following Mendel's laws. So that is very briefly one sentence description of that evidence and there are more, some of them we will see here too. So what he did is;

(Refer Slide Time: 04:49)

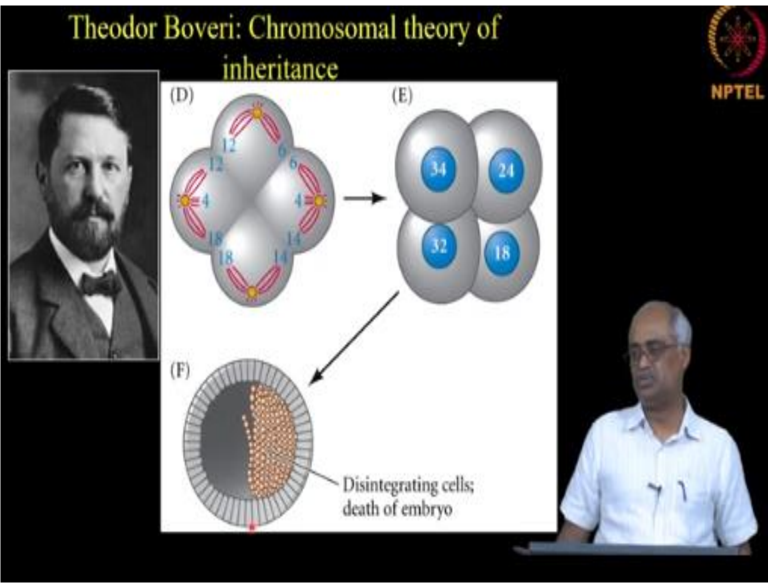


He took the Sea urchin, whose mode of reproduction is external fertilization. Many people used Sea urchin for studying fertilization, so that is another topic in the later part of this course. **(Refer Slide Time: 05:51)**



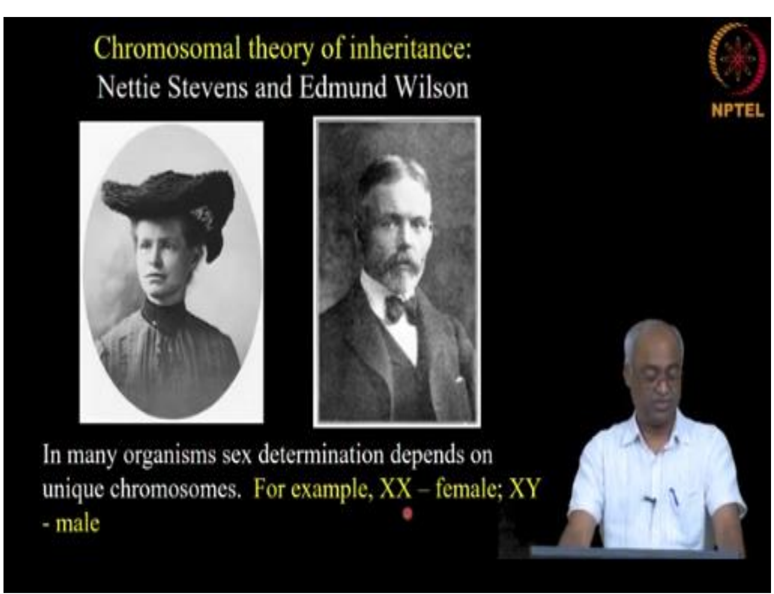
So he fertilized the oocyte with an excess amount of sperm such that he got some of the zygotes with more than one sperm or two sperm like here. Then he allowed the embryogenesis to progress, and he found the embryo setting up four poles like two spindles, and then it ends up making all kinds of weird spindles.

(Refer Slide Time: 06:21)



Then the embryo ended up dividing into four cells at a point where it should have divided into two cells each one having varying numbers of chromosomes. He isolated the blastomeres and showed that due to chromosomal abnormalities cell divison did not progress and cells died. So here you are seeing that cells disintegrating in an embryo and death of the embryo. So this is one of the evidence indicating that normal embryogenesis requires a normal number of chromosomes. So, therefore, chromosomes matter.

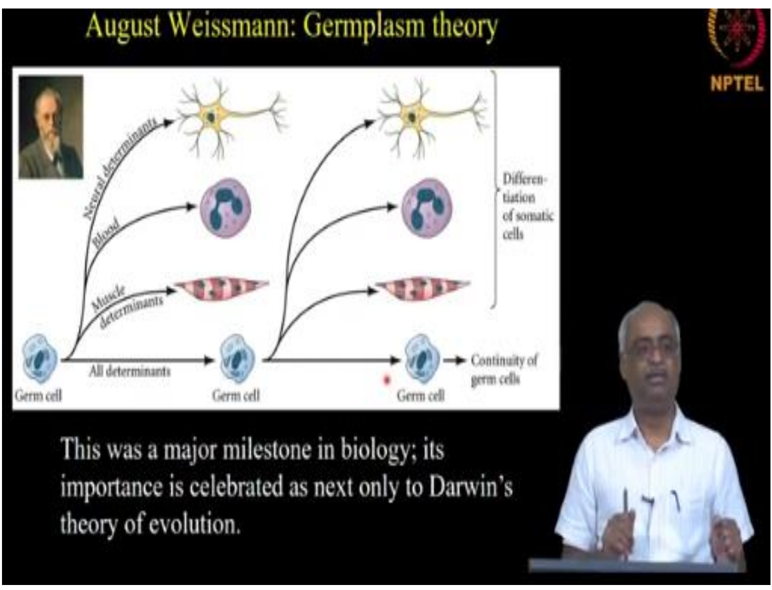
(Refer Slide Time: 06:55)



Nettie Stevens, was a postdoc of T.H. Morgan who is a very famous geneticist, but initially, he supported cytoplasmic control of development. Yes, So the embryologist was of two schools at that time. So Nettie Stevens strongly felt it is the chromosomes that controls development. She later worked with Edmund Wilson who also believed in it. They showed in many organisms sex determination is by the chromosomes. XY or XO in many organisms are male, and XX was female. So T.H. Morgan tested this more rigorously and nailed it down. He found that Drosophila eye color gene is inherited, on X chromosome, it is a sex-linked inheritance and that eventually proved that genes are on chromosomes.

So that is how science progresses; people do not get stuck on an idea forever when evidence surface they change their opinion. So this is some fascinating history to know. See, I am assuming some of you are going to be budding scientists, and therefore, the way thought process shaped up in the past scientists would be useful.

(Refer Slide Time: 08:34)



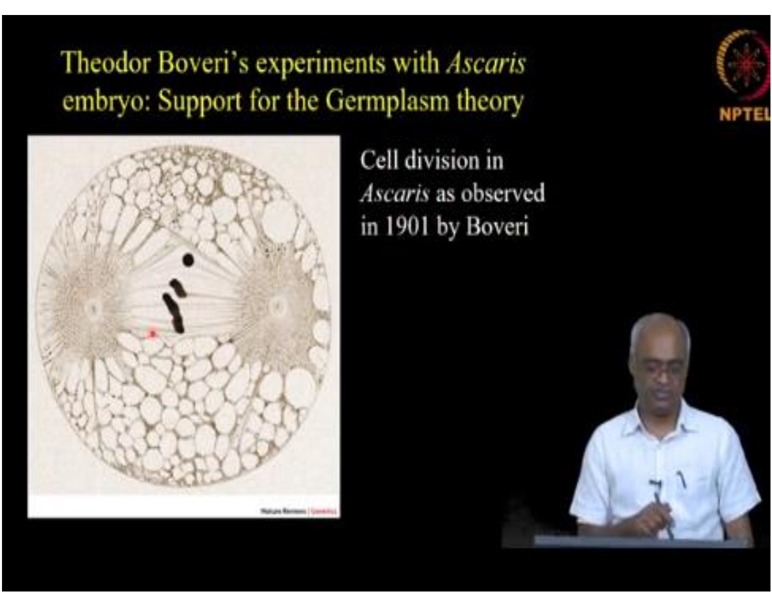
So all this evidence show chromosomes are essential, and but if all cells have the same chromosomes and chromosomes direct development, then how different kinds of cells arise, how come the same set of chromosomes make heart and lungs as well? So that is a big question. So August Weismann proposed a theory called Germplasm theory. So this is one of the critical milestones in biology, so many biologists consider this germ plasma theory and its importance as next to Darwin's theory of evolution. So what is this theory, he was the first one to distinguish somatic cells from germ cells. So he proposed that the germ cells maintain all the chromosomes, the genetic material and it is passed on from generation to generation; whereas the somatic cells that come out of these germ cells in every generation that is the gametes fuse to generate the zygote and from the zygote you get germ cells as well as the rest of the body. For example, muscles receive only the instructions needed to make muscle and blood cells receive only the instructions to make blood cells. Similarly, Neurons receive information to make neurons.

So this is what is germplasm theory is, and this is how he thought different cells become different kinds of cells. So Boveri's further experiment was a significant milestone that helps us to distinguish Lamarck's view of evolution from Darwin's view of evolution. So use-disuse theory would demand the information in the muscle go to the next generation. So that is the only way learned characters could be acquired, and germplasm theory explains why that would not happen because the information from muscle does not go to the next generation. So only the germ cell goes. So, therefore, random selection from changes that occur in the germ cell chromosomes is

what gets selected in every generation. So that is how this germplasm theory helped in understanding Darwin's theory of evolution.

The main component of Darwin's theory of evolution is not that he first told evolution happens; the main contribution is finding a mechanism for evolution. So the mechanism is natural selection, and for that, this theory was instrumental in explaining why acquired characters are not going to be passed on directly to the next generation.

(Refer Slide Time: 11:50)

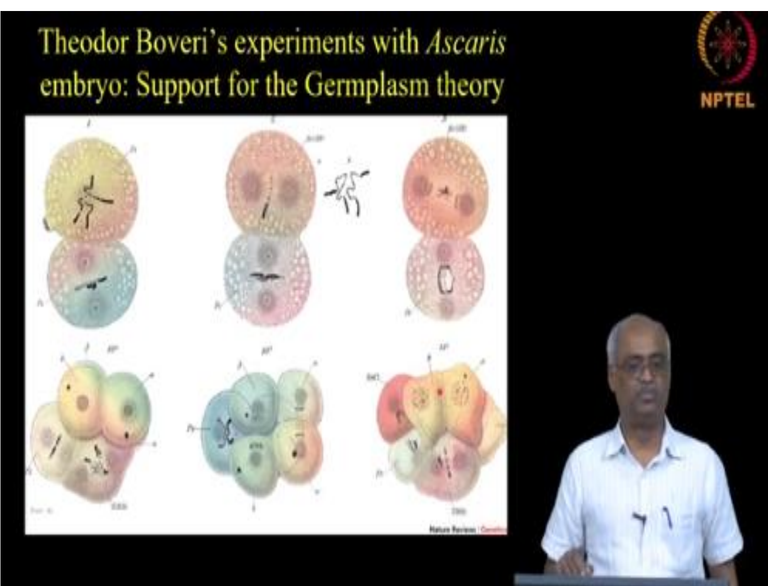


So Boveri did additional experiments that sort of supported germplasm theory. That is primarily because of the organism he chose, so *C. elegans* was not the first nematode that was used for studies. In 1901, Boveri used Ascaris, which was a parasitic nematode. Boveri died of infection by Ascaris years later. So the reason he chose Ascaris is it has only two chromosomes, so for addressing specific biological questions, you need to have knowledge of organisms. So only then we will know which organism is suitable to address a particular question without which you would just not be able to answer. So that is why it is more essential to learn botany and zoology before actually, you learn molecular biology and biochemistry.

So that is why we will not find *C. elegans* or Drosophila or Arabidopsis because we do not learn organisms. So he knew that Ascaris has only two chromosomes, and therefore, it will be easy to

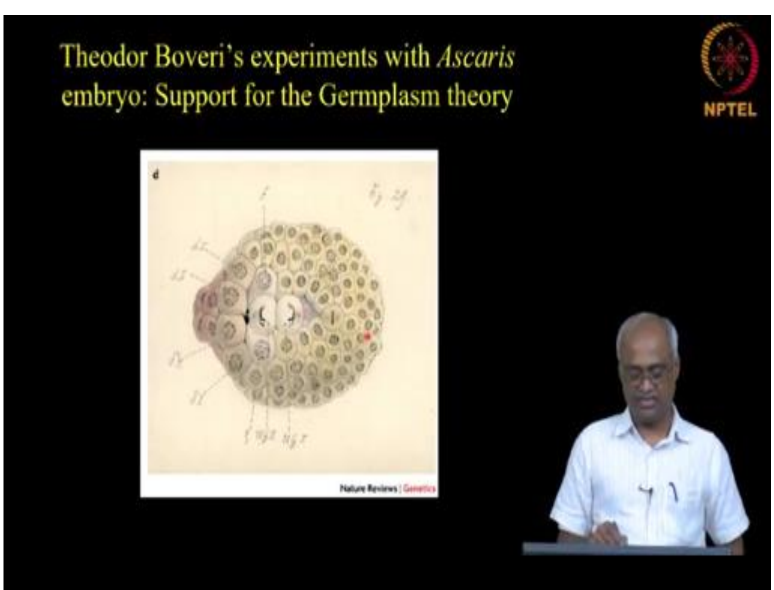
observe what happens to these two chromosomes as cells divide. August Weismann's theory proposes that the somatic cells will have different contents.

(Refer Slide Time: 13:24)



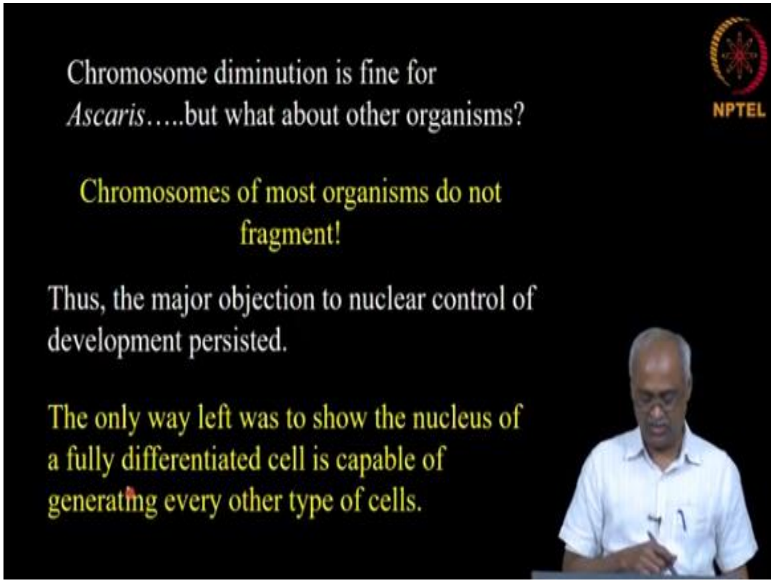
Boveri followed the cell division of Ascaris embryo; he used the names that are used in *C. elegans* like EMS, endoderm muscles come from that, P1, P2, P3, P4 are the germline lineage. So here if you look at this, in the first diagram the bottom one is going to be the future germline. So you have the chromosomes there, and if you go to the second diagram, the top cell is setting up the next spindle for the next division you see the chromosomes are breaking in, so this is what is happening in other cells. In contrast, the cells that are going to be future germline remains intact. You go through this over division, so chromosomes in the ones that are going to be germ cells are intact, but in the adjacent cells, which are the somatic blastomeres, the chromosomes are fragmented. So this process is called chromosome diminution.

(Refer Slide Time: 14:39)



And you go further enough embryo where the cleavage is over the two primordial germ cells retain the chromosomes intact, while in others it has broken into smaller pieces so this is a strong support for germplasm theory.

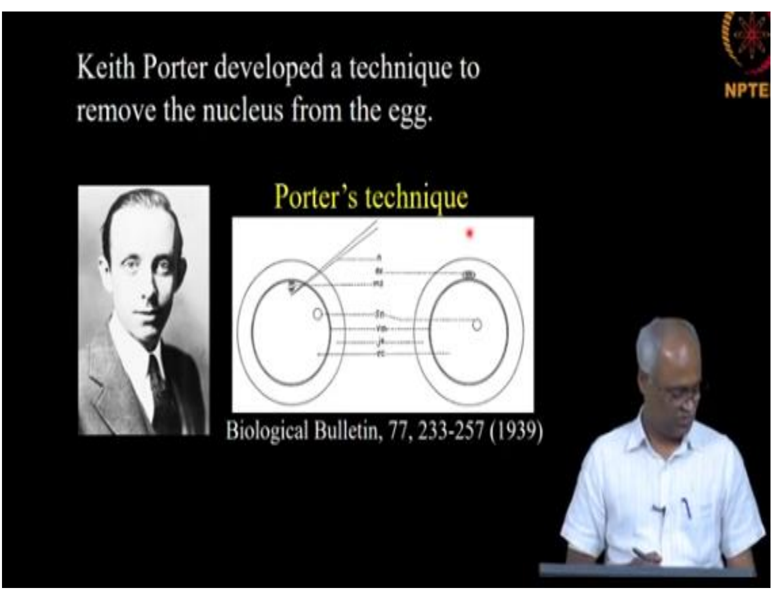
(Refer Slide Time: 15:00)



So that is Ascaris. But is this happening in every organism? No. So that means there must be other explanations in other organisms to support germ plasm theory or germ plasm theory may be wrong. So, therefore, the objection for nuclear control of development persisted. So people still thought that nucleus may not be everything, meaning chromosomes may not be everything; cytoplasm probably still is the key.

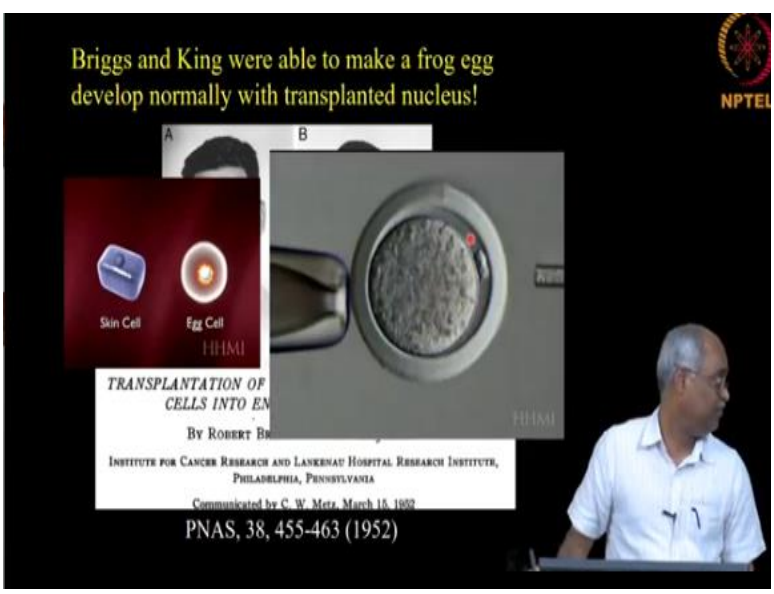
So the only way to address this is to take the nucleus of a fully differentiated somatic cell and show that it can direct an egg into a fully developed embryo. So that required somatic nuclear transfer. That needed some technical advancement where nucleus can be removed out of an egg and, at the same time, activate the egg; cytoplasm gets activated when sperm enters. Then you introduce a somatic nucleus, and so this had to be figured out, and people did that eventually.

(Refer Slide Time: 16:18)



Keith Porter developed a technique for doing this. So he found that by poking in and trying just to move the nucleus out, not only was he able to remove the nucleus, it also ended up activating the cytoplasm of the egg as well. So this is called Porter's technique. So he found a way to remove the nucleus from the egg.

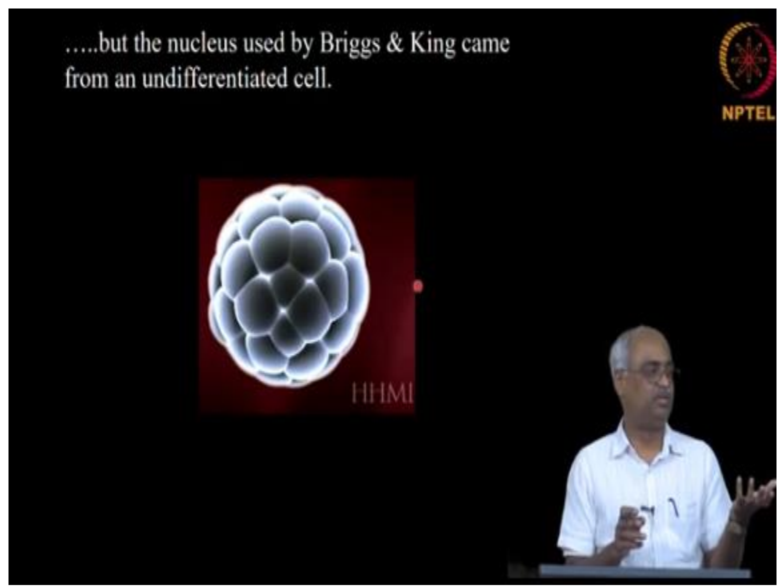
(Refer Slide Time: 16:42)



And making further improvements on that, Briggs and King, they were able to show that transplanted nucleus from other cells can direct a frog egg to develop normally. But the only difference is that they were able to go up to a certain stage like up to blastula stage but not up to the tadpole stage and that wait needed some more optimization, and they had to work on a different set of frogs.

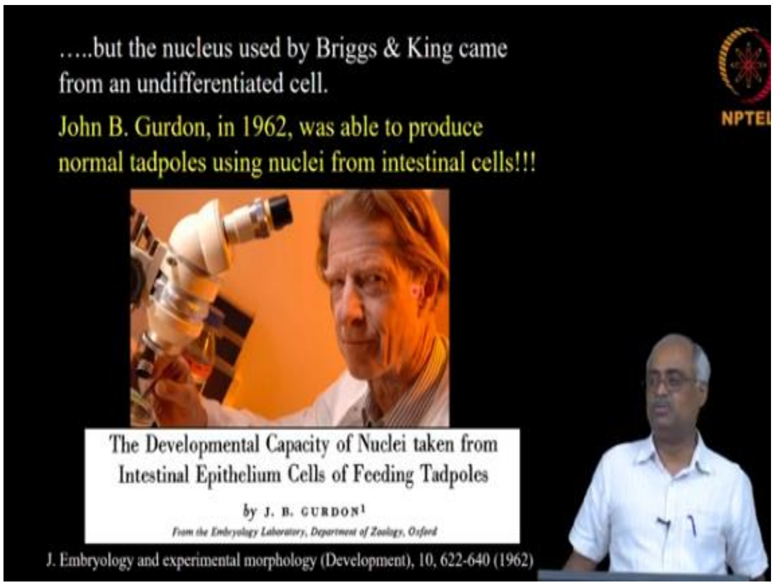
And so here is an example of how this was done. So this is a movie, but I am sure the movie is not going to play. So essentially, you have a skin cell and an egg cell, so you take the nucleus out of the egg cell and then you introduce the nucleus from the skin cell and then allow it to develop. The way you do that is, so this is an egg. Here you have a capillary where you apply gentle pressure; therefore it is held in position. It does not move away. So this technique will vary from organism to organism. This holds true for most of the mammals and even in vertebrates. So here is the suction applied. So it is held in place. So then this is the needle that you are going to poke in, and in that, you are going to take the nucleus out gently and the other nucleus that you get from the skin cell is put into this and then its nucleus is going to be diploid. You do not have sperm in this and you watch what happens.

(Refer Time Slide: 18:33)



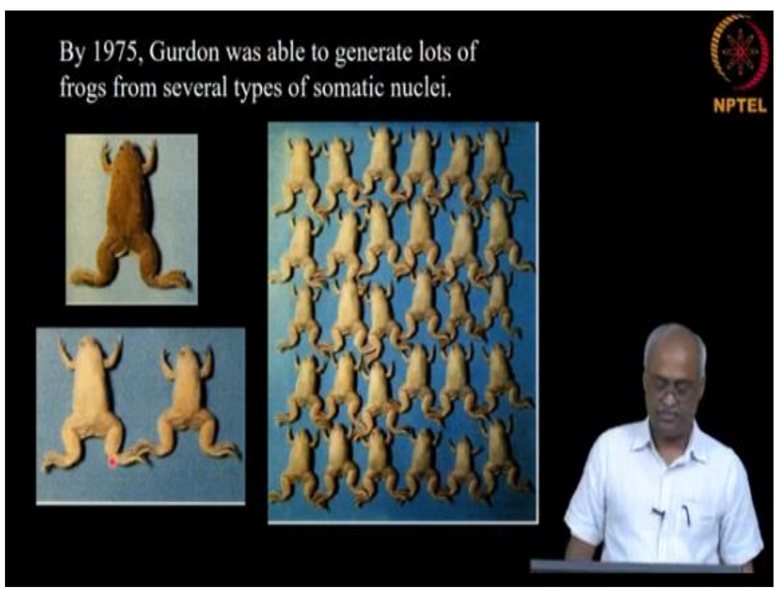
So there were two issues with Briggs and King Experiment. One is that it did not go beyond the blastula stage, and the other one is they used nucleus from another embryonic cell. It is still an undifferentiated cell; it is not equivalent to a fully developed somatic organ.

(Refer Time Slide: 18:50)



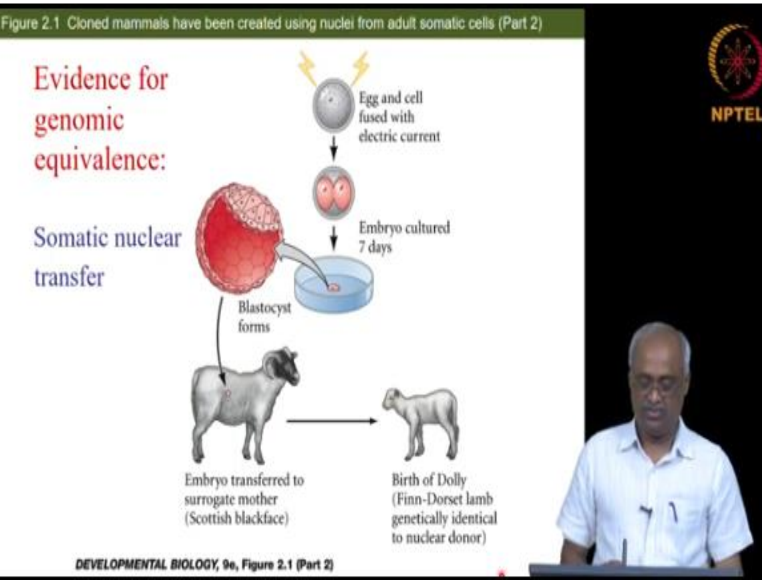
So John Gurdon, who got the Nobel Prize just a few years ago, less than ten years for the work done in 1962. So he got a Nobel Prize along with the Yamanaka group for doing somatic cloning in the mouse. But they did it very recently in mouse, but Gurdon had done in the 1960s, and he showed that you could take intestinal cell nuclei and introduce by using Briggs and King's technique, and they can develop it up to tadpoles.

(Refer Time Slide: 19:36)



And using different genetic makeup, like for example, here you have a dark-colored frog, and you take the nuclei from these light-colored ones and using its egg, then you show that all the progeny are like the nucleus donor. Showing that, two things one nucleus directs the development and other evidence for genomic equivalence. He is taking intestinal cells, and they have it. Now of course, people have done with other somatic cells as well.

(Refer Time Slide: 20:16)



This is one more evidence, so I am sure some of you were born by the time this was published, it is that recent. So Wilmut in Scotland and his group were able to do a very similar experiment in sheep. From the oocyte donor's egg, the nucleus and the spindle were removed, and then the udder

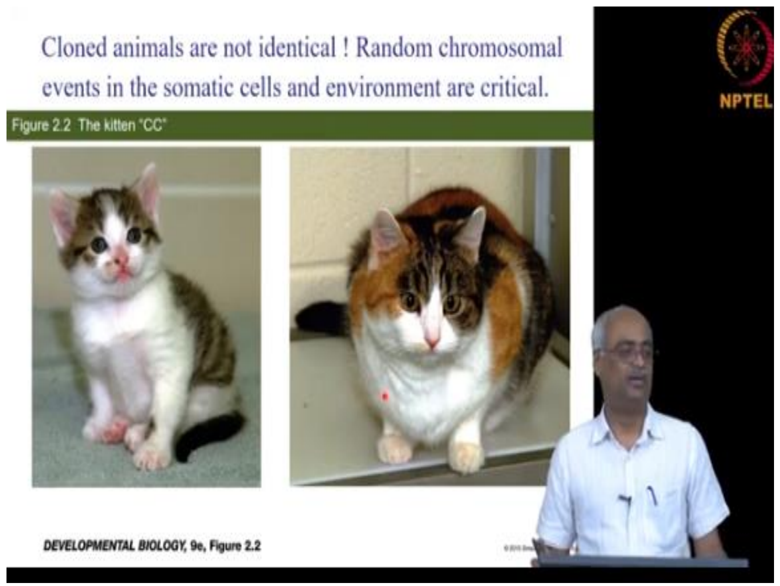
cells from the nucleus donor were transferred into the enucleated oocyte, and then it was allowed to develop.

So electric shock allows the membrane-membrane fusion, then you culture the embryo up to blastocyst stage, and then implant in the surrogate mother and then the progeny is genetically identical to the nucleus donor, not the one that provided the oocyte. So this nails down that there is genomic equivalence, as well as the nucleus drives the development.

(Refer Slide Time: 21:43)

So this is the Dolly and its baby here. But this is not to say that DNA alone is responsible for development. So there are variations to this mostly true concept. Those variations are subtle variations but still important.

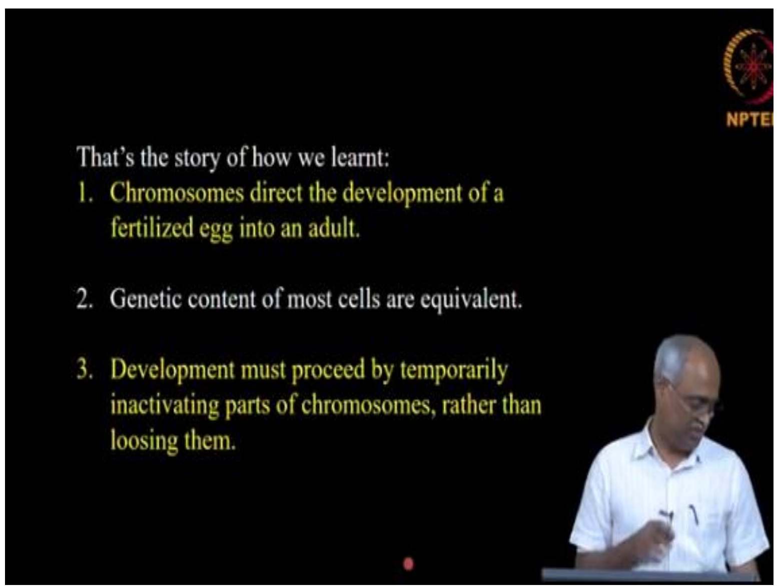
(Refer Slide Time: 22:08)



One of them is illustrated here. So this kitten is a somatic clone from this cat, and in these cats, the coat color has random variations, and that is because of random inactivation of one of the X chromosomes. But to get to that, there is a concept called dosage compensation. So all the girls have two X chromosomes, and boys have only one X chromosome. So are we going to have double the amount of X chromosome output or are we going to have half of the output in boys? But in the end, it is the same output that is because the one extra X chromosome is inactivated. So there are multiple mechanisms to handle the dosage compensation, but we will not get into all of them. We will just consider the organisms where one X chromosome was inactivated. So which X chromosome is inactivated and at what stage in development. So that varies from organism to organism, and in this cat, it is random inactivation in every one of the somatic cells. So depending on in which part of the skin which X chromosome is inactivated you get different color patterns and due to that this kitten does not look identical. So that is one, and these are epigenetic modifications. The DNA remains the same, and DNA is responsible mainly and there are other situations like the environment that can still have influence.

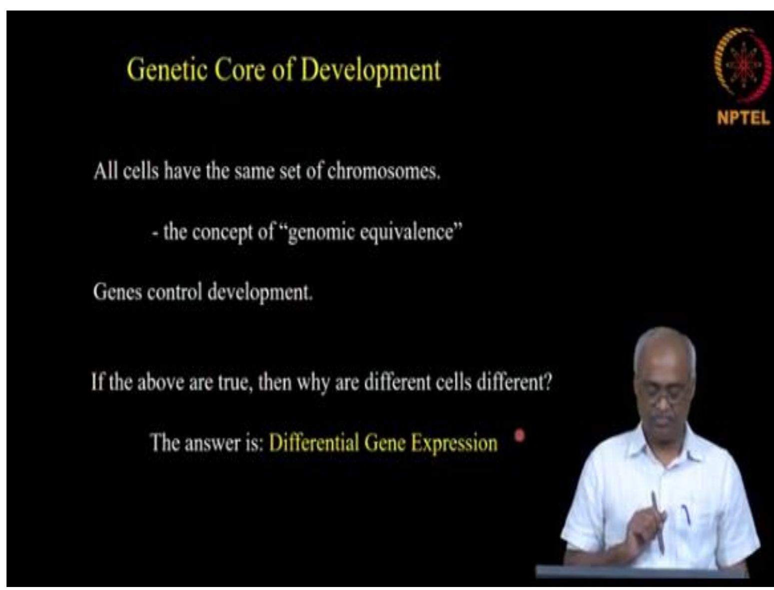
So DNA is not the sole determinant, but for the original question, the answer is yes, it is the set of chromosomes that determine development.

(Refer Slide Time: 24:03)



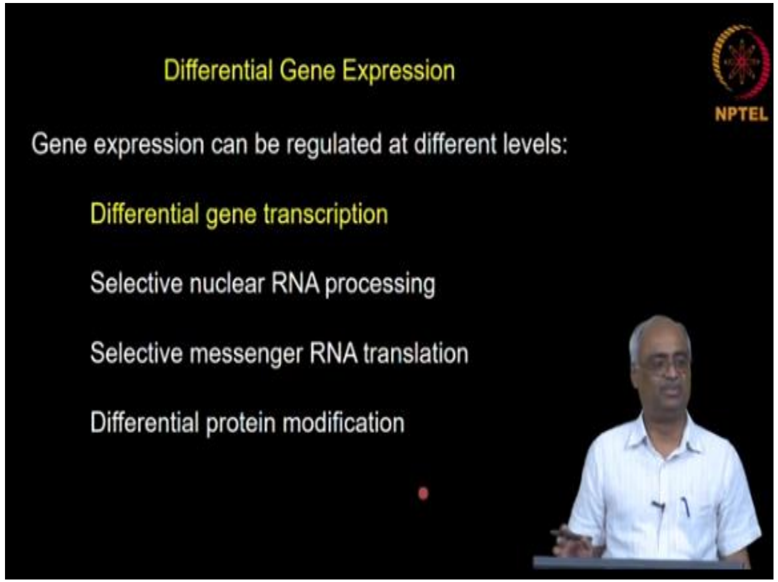
So this is how we learned the chromosomes direct development, and the genetic content of most cells is equivalent. I am saying most because there are variations like our RBCs, immune cells. See if that is the case, then the same set of chromosomes, how they direct development? Each cell has to become a different kind of cell; how does that happen? Therefore the postulate here is development must proceed by temporarily inactivating parts of chromosomes. Because the chromosomes still exist, there is no chromosome diminution, as we saw in Ascaris. So that means part of the chromosome is expressed in one type of cell and another part is expressed in another kind of cell and so on and that is what brings us to a concept called differential gene expression. So most of the modern developmental biologists deal with this idea of differential gene expression. So that is what people focus on, and much of the research is about that.

(Refer Slide Time: 25:09)



So genes control development, and if it is true what is the answer, it is the differential gene expression. So that is the end of this discussion. So we will move on to our next thing, we will look at how the differential gene expression happens.

(Refer Slide Time: 25:32)



So gene expression actually can be controlled at different levels. I am sure many of you may be already familiar from your molecular biology class, but to give continuity, I will quickly go through them. So, you can have differences in gene transcription; for example, the globin gene is not transcribed in pancreatic cells, which produce insulin, and other cells do not produce them. So that is a transcription level regulation.

A gene may be transcribed in one type of cell but not in another type. So that is differential gene transcription, and second, a similar difference may exist at the level of nuclear RNA processing, for example, splicing and export out of the nucleus might vary. In a given tissue, a particular gene's mRNA may not be exported while the same mRNA is exported into the cytoplasm in another type of cell. So, that is selective nuclear RNA processing, and then you have selective messenger RNA translation. For example, if you take two tissues like germ cells or neurons, there the response required is rapid like if you take neurons; for example, the response required is swift. So you cannot be activating the expression of a new gene all the way starting from transcription onwards, you may not have that time. In such situations, you have mRNAs already made but not translated until it is needed. Similarly, if you take the biological process of reproduction, early embryonic development, during the rapid cleavage, will not have the ability to do everything by starting from transcription onwards.

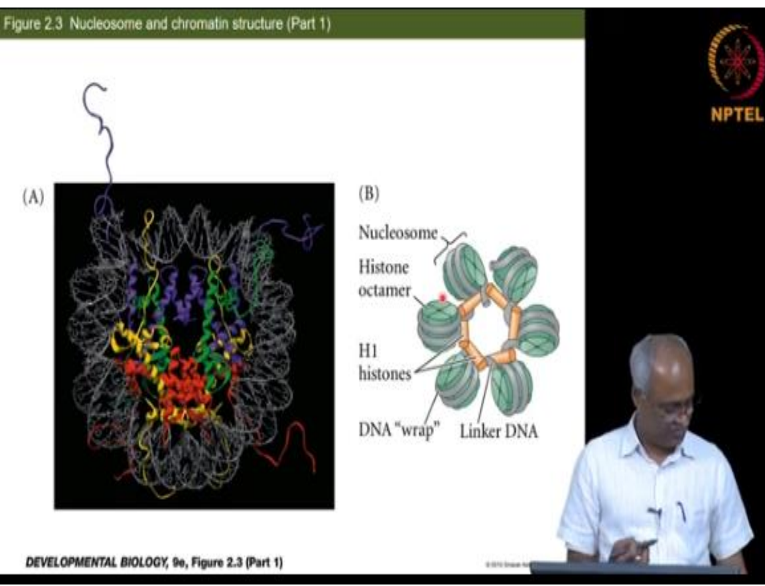
A lot of proteins are required during early embryogenesis and their mRNAs are already transcribed in the mother's germline and brought in via the oocyte cytoplasm where the mRNA are kept translationally quiescent. Their translation is sequentially activated as and when they are required. So these are two perfect examples where translation control plays a significant role, so that is how you have selective messenger RNA translation.

That is another step of gene expression where you can have control to bring about differential gene expression and another important step. So these are the four key steps, but there are other sub-steps in each one of these where you can have gene expression control. Do not think gene expression control means transcriptional control that is a common misconception among many students who have not studied developmental biology.

But after this course, you are not going to have that feeling, and you can also have at the level of protein modification. So you have seen two kinds of protein modification if you have studied in the biochemistry class one is zymogen activation like protein is made as pre, pro-protein. Then it is cleaved to generate the final product like some of the hormones and blood clotting factors and so on. Then you have other proteins where post-translational modifications such as phosphorylation activate or inactivate proteins.

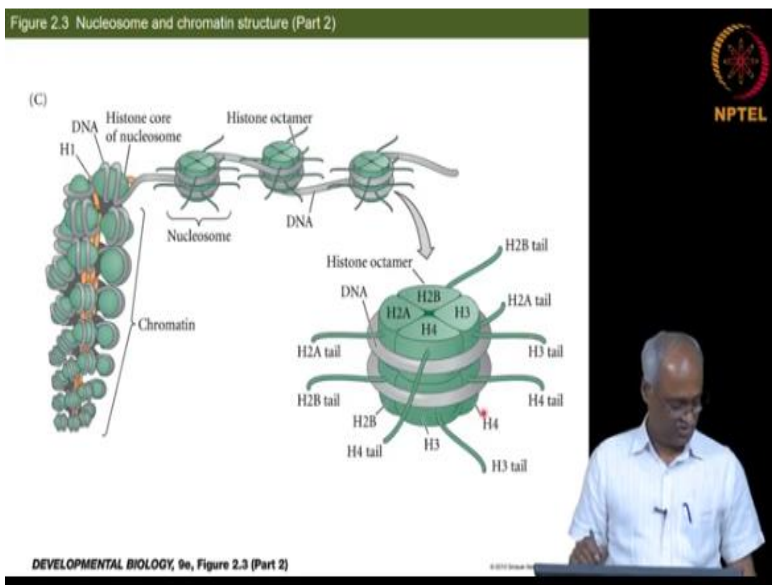
So these are the levels at which you can make differences among the cell types in terms of what genes are expressed and are not expressed. We are going to look at these regulations, and before we go into this, we are going to have a good idea of chromosome structure and the eukaryotic gene structure. That is what we are going to do in the next few slides.

(Refer Slide Time: 29:44)



So I am sure you are familiar, but I want to go through this quickly. So this is a crystal structure of the same here. It is the chromosome with all the proteins; proteins are histones here. So you have the different histones, the four different ones differently colored here and the DNA is wrapped around it and pay attention to these tails, so these are histone tails coming out. So they are essential for our discussion, and you have one Histone, this orange rod-like structure, H1, that plays a crucial role in really compacting this structure into a long spring-like structure.

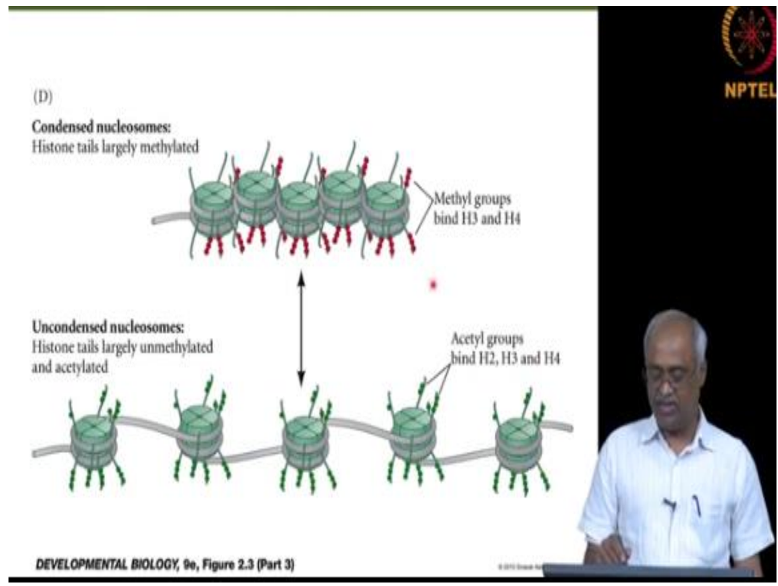
(Refer Slide Time: 30:37)



I am particularly and intentionally avoiding the word solenoid because some of you will have to go and look at the dictionary to find what solenoid means it means spring, a coil. So this coiling is possible because the H1 that binds here, that can pull them together and make them become a coil like this. So this is how our chromosome exists, as a tightly coiled condensed structure.

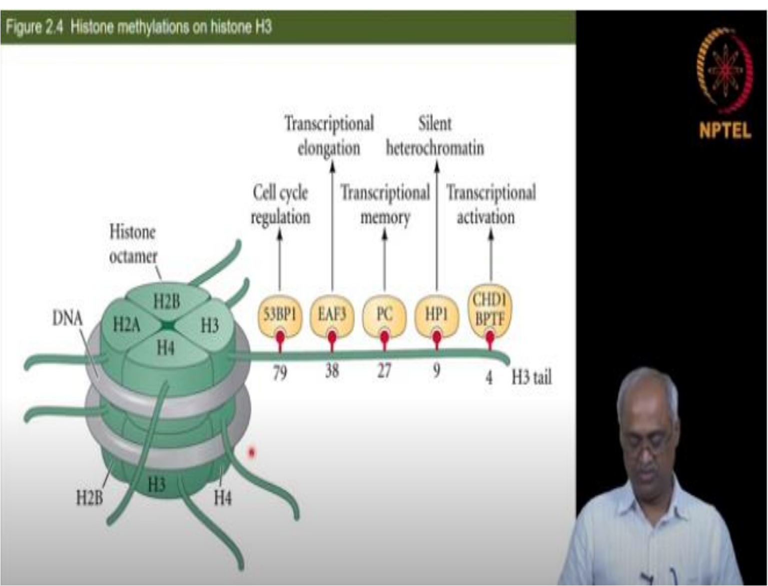
So it is supercoiling because here itself you have coiling. Then you have an additional layer of coiling, and each one is the nucleosome, so they contain eight molecules of Histones, H2A, H2B four of them, and then H3 and H4 each two. So that is the structure here, and this is how a chromosome exists. Now let us pay attention to these tails.

(Refer Slide Time: 31:41)



So some parts of the chromosome are tightly condensed, and we call them heterochromatin, and there most of these tails on H3 and H4 are methylated. The methylation commonly promotes this coiling, and these are usually transcriptionally silent and are not accessible to the transcription factors and RNA polymerase. Unless otherwise, a special kind of transcription factor called pioneer transcription factors that we will learn later. In uncondensed nucleosomes, the Histones lack the methylation marks, not all of the methylation marks but majority of these and they are usually acetylated. So acetylation in H2, H3, H4 marks active chromatin. So here, you see control of gene expression at the anatomy of the chromosome. And it is mostly regulated by modifications to these tails, and that is why I was telling pay attention to the tail at the crystal structure level itself. So this gives you an idea of how they are available readily or rather readily accessible for enzymes that could add or remove methyl group or acetyl group.

(Refer Slide Time: 33:22)

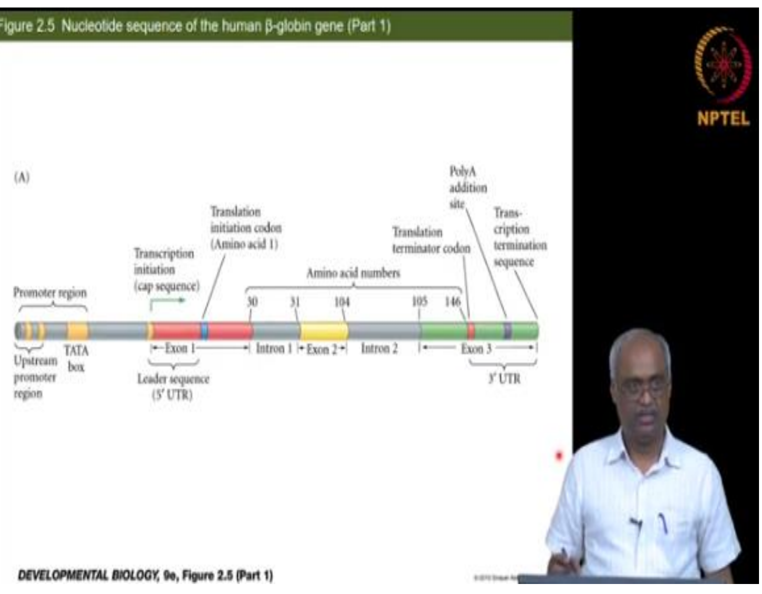


So, now a closer look at one of the nucleosomes mainly focusing on H3 tail. So here if you look at you have these red ones which are the methylation marks. So some of the red methylation marks combined with acetylation, in general, activate transcription. That is the point I want to make here. It is not that methylation means inactivation regardless of where the methylation is, so here these numbers 79, 38, 27, 9, 4 etc. refer to lysine residues in H3 starting from its N terminus to C terminus.

So, depending on which one of these lysine's are methylated and whether the overall chromatin is acetylated or not, determines whether it is going to be active or not. These are shown here, for example, when you have H3K4, H3's lysine 4 is methylated then this factor is going to bind, and that leads to transcriptional activation, and here H3K9 methylation silents heterochromatin. So this is how modifications to Histones can have an impact on gene expression.

So here, the control is at the transcriptional level. So the chromosome anatomy itself affects transcription, for example, addition to methylation, which is a post-translational modification of the Histone, which is a protein, you have methylation of nitrogenous bases on DNA as well, for example, cytosine methylation. So do not get confused between the two methylations. Thus in a chromosome, DNA can be methylated, which is not our current discussion, and you can have proteins that are methylated as well. Here we are talking about protein methylation; the protein is Histone.

(Refer Slide Time: 35:32)



Having looked at the chromosome as a whole now we will refresh our memory about the structure of a eukaryotic gene. I do not understand what the problem is, but most students after going through molecular biology course even after a couple of years later when you ask them to draw the structure of a eukaryotic chromosome gene they have problems and people do not understand, for example, where is a promoter with respect to start codon and what is a 5' untranslated region (5'UTR).

Is 3' untranslated region (3'UTR) is present in the chromosome or not and is the Poly A tail present in the chromosome or where does it come from. So like that people have confusion, to avoid that we are going to refresh our memory of the structure of a eukaryotic gene. So this horizontal bar represents the DNA sequence of a part of the eukaryotic chromosome. So here, first we will focus on what is easy for us to understand that is the coding sequence.

So the coding sequence is divided into exons meaning there are intervening sequences that are not going to be part of the matured mRNA. So there are intron 1, intron 2 here which needs to be removed in the final mature mRNA and have the sequences corresponding to this exon 1 in the DNA, exon 2 and exon 3 right after that. Then starting from the start codon, which is ATG coding for methionine, you go all the way to stop codon.

When you look at the gene structure, these coding sequences are divided into exons with the intervening introns. Then when you look at the 3' end you have further sequences, and they are called the 3' untranslated regions, meaning these are present in mature mRNA, but they do not code for amino acids.

These are sequences beyond stop codon, so the primary point I want to emphasize is the mRNA has sequences beyond the stop codon, and that is called 3' untranslated region because in terms of the directionality of the mRNA it is 5' to 3' so it is 3' untranslated region. So that is coming from the chromosome; it is transcribed from the chromosome and it is retained in the mRNA therefore it is an exon. 3'UTR corresponds to an exon just like the amino acid coding sequences and in that RNA sequence, 3' UTR have sequences like poly-A, polyadenylation signaling sequence which signals two things, one cleavage of the RNA from the primary transcript and then promoting polyadenylation, like poly A tails are added. Some amount of poly-A tail is added in the nucleus, and further extension of that happens in the cytoplasm.

So poly-A tail is not present in the chromosome, it is a separate enzyme that adds multiple A continuously one after the other. It does not require a template because you are continually adding A to any existing RNA. So that is how that is made. The summary is that after stop codon you do have RNA sequence and that is called 3' untranslated region, and that contains a polyadenylation

sequence that helps in cleavage of the primary transcript at the 3' end addition of a poly-A tail and let us go to the 5' now so.

In the 5' similarly before the ATG you have sequence and that is called 5' untranslated region and that usually starts with the base called transcription start site and that is modified like the 3'modified with poly A tail, this gets a cap, a 5' cap it is a trimethyl G added it to it. So that is the RNA sequence, but here in the DNA sequence upstream of the ATG, you are going to have some RNA sequence.

That is also part of the exon o in the genome while the trimethyl G is not part of the genome. So then upstream of that, you have a promoter. The promoter is where RNA polymerase is going to bind and start transcribing. The first base that is transcribed is the transcription start site or the initiation site that is not ATG, ATG is going to come further downstream at the end of 5' untranslated region and the promoter region can have multiple parts in it which we will discuss in detail as we go.

TATA box is one that is present in all promoters. So these are like core promoters, the promoter sequence which is required for transcription of any gene and particular sequences help in differential gene expression so that we will talk about it when we go further.