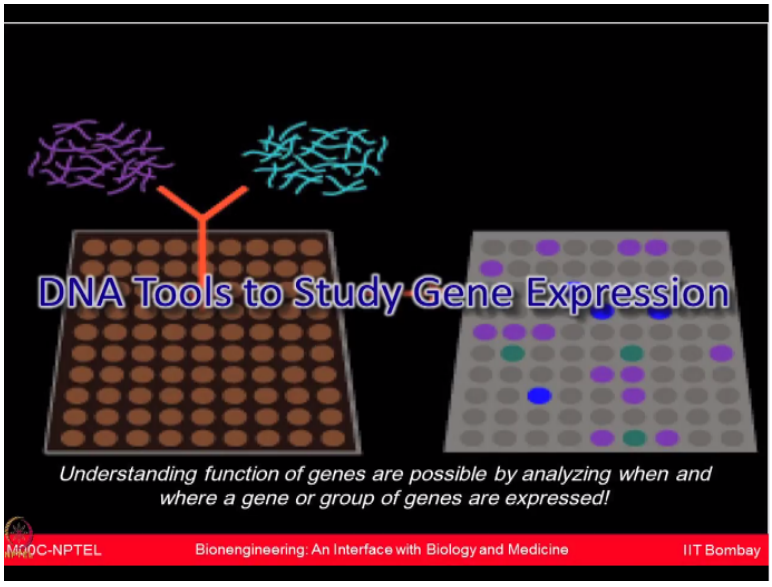
Bioengineering: An Interface with Biology and Medicine
Prof. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology – Bombay

Lecture - 12
DNA Tools & Biotechnology-IV

Welcome to the MOOC NPTEL course on bioengineering an interface with biology and medicine. In today's lecture, we are going to continue learning several biotechnology tools that can be used for research, clinical applications and medicine. We will learn how to study gene expression. I would like to make you familiar with new tools like cDNA synthesis and reverse transcription PCR or RT-PCR.

We will also briefly talk about a promising technology which is microarrays which has made huge impact in the genomics field and how even the protein community has started making use of the microarrays for various application including expression analysis. Let us start today's lecture now.

(Refer Slide Time: 01:05)



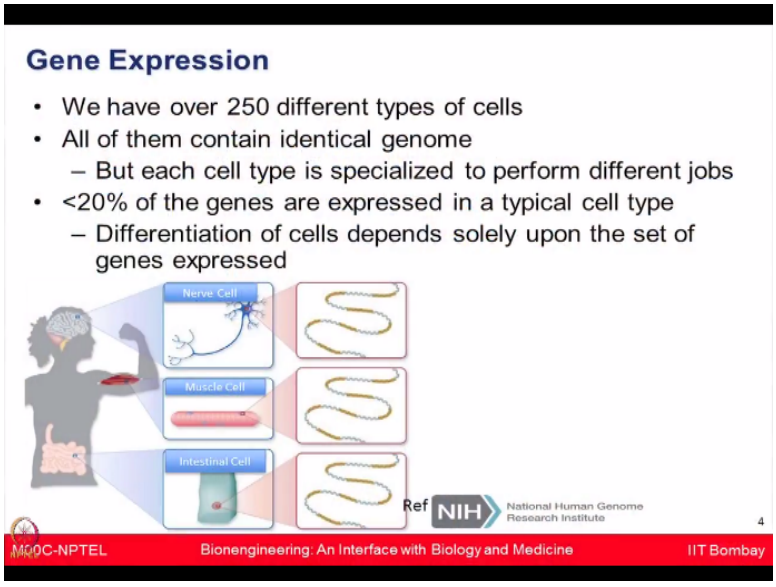
If you want to study which gene is expressed at which time, for example we say you know somebody the normal individual and one person is suffering from a given cancer type, now which are the proteins or the genes expressed in this cancer individual? How to study that? So you want to know that you know which are the mRNA being expressed, how they are being different from the normal individual, which are the proteins being made, how they are different from the normal individual.

So this is known as expression analysis. You are looking at from the normal basal level how an mRNA or the protein is going either upregulated or downregulated. So this is the kind of change which you want to study that you are studying as a part of the expression analysis. So understanding the functions of genes are now possible because one could analyze when and where a group or group of the genes are expressed.

Think about a broader picture, so we have millions of cells right and in each cell ideally we should have the same genome because if you look at the macro picture of the cell which we talk that we have the you know the nucleus and all the genes present over there. So now ideally all the cells will have the identical genome but each cell type is having certain specialized role to play.

So for example a cell from pancreas versus a cell from neuron or a cell from heart is going to be slightly different and they have to play very specialized role.

(Refer Slide Time: 02:32)



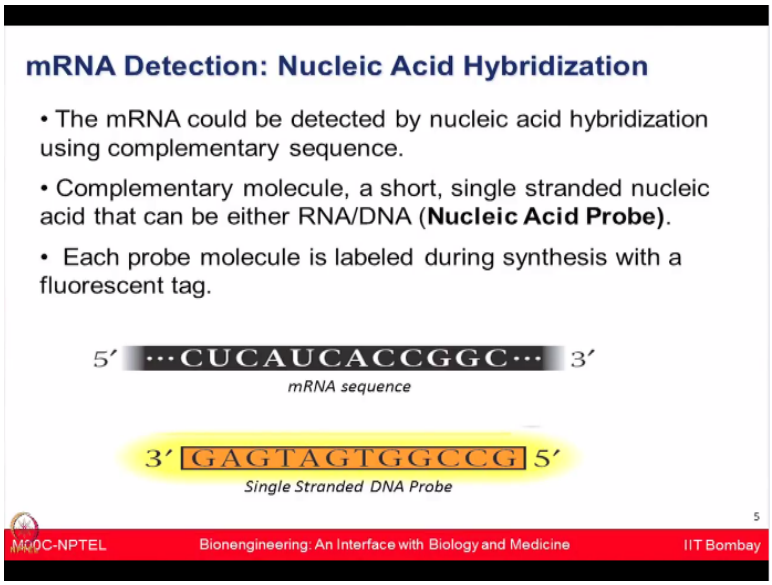
So what it shows this particular image which is a reference from National Institute of Health, NIH that we have all these genes and probably you cannot read clearly. There are certain introns and exons and some genes are on and off. So in these cases all the genes are same whether you talk about a nerve cell, you talk about a muscle cell, you talk about intestinal cell, all of these cells having the same genes.

But just imagine when you need specific requirements for certain proteins to be on then you will see the expression of some genes. So only a few genes are expressed in our cell and some different genes are expressed in muscle cells, not the same genes are expressed everywhere all the time because they do not want to use all its energy to just express everything everywhere.

So specialized genes are being expressed in a specialized cell type and that is what is shown you here that while all the genes are present all the time but only specialized genes are being expressed in specialized cells and that is going to give the unique ability for those organs and those particular functions to tackle. Now let us think about you are from a central dogma point of view, we have from DNA to RNA and then proteins being made.

So the mRNA which is the first functional information which comes out from the static genome, if you want to derive that sequence.

(Refer Slide Time: 03:52)



So let us say this is the DNA sequence of you know given gene of interest we have. What will be the RNA sequence from it? So you are just making everything complementary, just keep one thing in mind that rather than having T you have the U here, you have uracil right. So if you had to make the complementary sequence of this DNA and convert as mRNA which I think you should try out and this is very simple exercise for you to do.

It is like you are just making the complementary sequence but you are replacing the T's with U's right. So it means if you want to study what is the mRNA synthesized for a given gene,

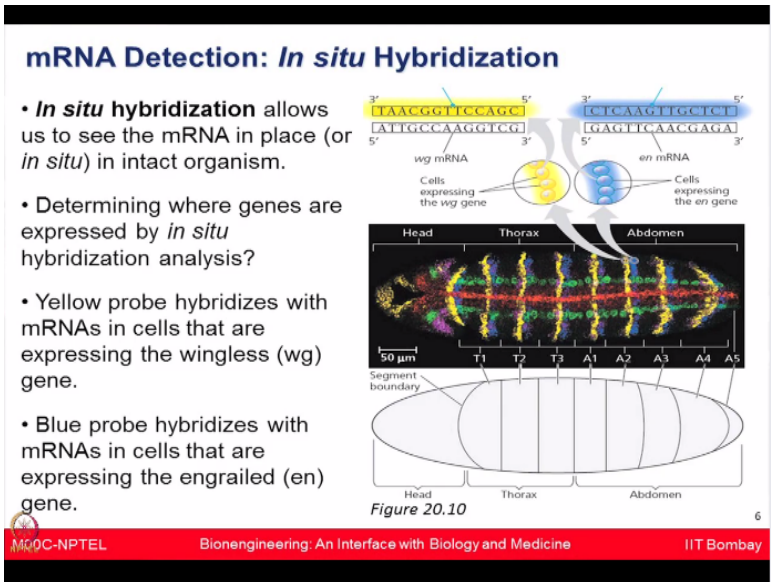
so you know the gene sequence and now you can make a probe at the mRNA level and that probe is like a chemical and that chemical will have this sequence which is not having T's but rather having U's alright.

So the mRNA could be detected by a nucleic acid hybridization technology which is just complementary in nature. You are making the DNA strand and complementary mRNA sequence from it. This complementary molecule which are usually short, single-stranded nucleic acid molecules can be either DNA or RNA. So whether you know this sequence you can design the complementary DNA probe or if you know this sequence you can design the complementary other sequence right, both way you can design.

So each probe molecule can be labeled during the synthesis process you know using some fluorescence which can be used to monitor where this particular gene is expressed. So you want to study that you know gene X is expressed in nerve cell or muscle cell and now to do that you have made this particular probe and this probe is having a label which is fluorescently labeled.

So wherever it goes that particular probe is going to bind and if you detect in the fluorescent microscopes or the scanners then you can see where this particular gene is expressed.

(Refer Slide Time: 05:51)



Another technique that people want to see in the life condition in the in-situ manner that where this particular RNA is expressed. This technique is known as in-situ hybridization. So this allows not only to know that you know mRNA being expressed or not but also it allows

us to see where mRNA being expressed in which cell and again people have used the model systems like Drosophila which can be used to do the experiments very easily and you can do lot of manipulation over there.

So in this case you know what you see here a Drosophila the structure and lot of these interesting colors which you see are actually from different fluorescence patterns which people have used as the labels for labeling different type of RNA. So the experiment objective here was to determine where the genes are expressed by doing the in-situ hybridization analysis.

To do this part they have used different probes. So when I say probe, it means you have synthesized, if you had known the mRNA sequence of a given gene then you have made a complementary DNA probe from it and that you have labeled with certain fluorescence tags. So in this case let us say we have made this probe which is for a gene which is wingless gene wg.

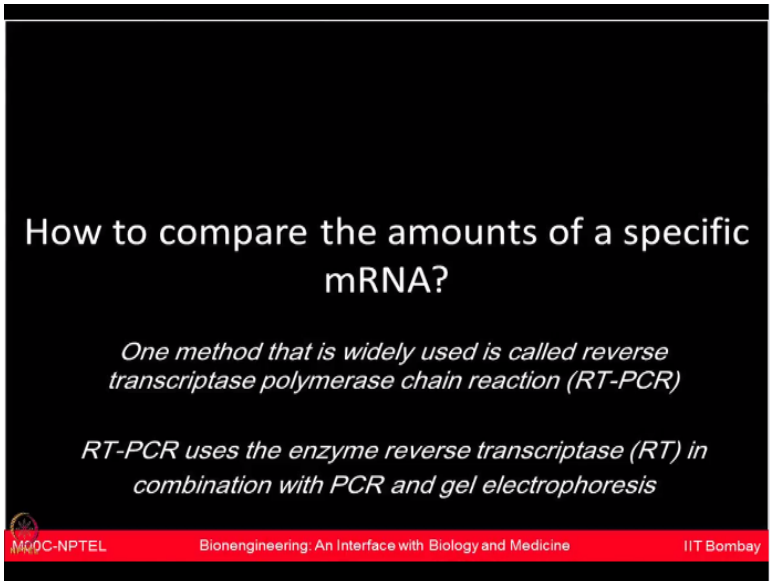
So this gene has property to make a wing or not, not make that wing and again that if you play with it you can make the wild type or the mutant forms of these genes right. So you have made one probe which is with yellow color which can hybridize with the mRNA and that can show you where these particular yellow genes are being expressed which is for the wingless property.

And you have also made another probe DNA probe which is a blue probe which hybridizes with the mRNA cells which are for the engrailed or en gene. So now by using different type of probes for the yellow color and for the blue color, you have made these probes and you can see that where these particular mRNA is being expressed in the cell. So it is very interesting experimenting.

Because you can see in the live conditions in the whole of the fruit fly where the gene of your interest is being expressed but just by playing with different colors. Just imagine how much you know progress we have made in knowing these sequences and these technologies that not only you can determine that you know a gene being expressed or not but also you can see in the live condition where being expressed and you can use different colors to actually show the different patterns.

So now if your interest is to look at large number of mRNA comparison as I said from a disease conditions versus healthy individuals, there are many methods which are being used in addition to what we just talked.

(Refer Slide Time: 08:37)

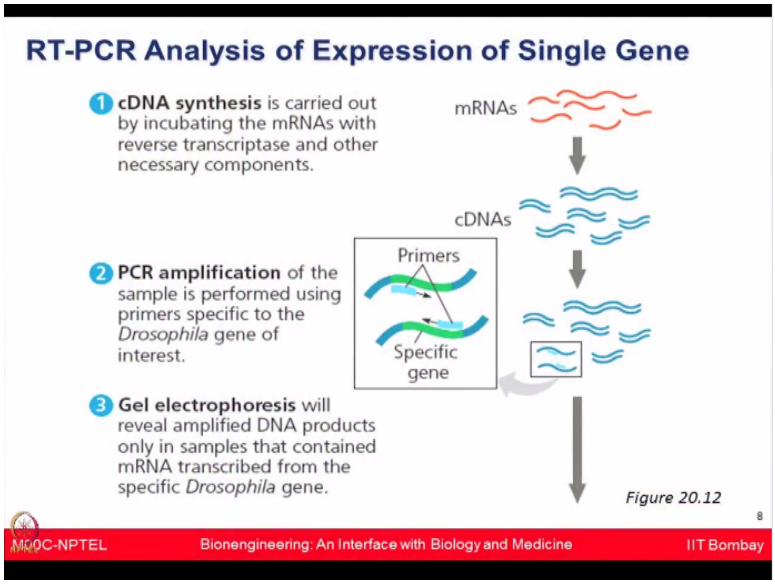


One method is also based on the polymerase chain reaction but that method is reverse transcription polymerase chain reaction which is RT-PCR. If you just think about the central dogma from the gene sequence to mRNA that process is known as transcription and from that mRNA to the protein is translation. Now just imagine the reverse of it from the mRNA you want to make the gene or the DNA.

So then what will be that process known as reverse transcription because you are not doing the transcription in the forward manner but reverse transcription is happening right. So people have used reverse transcriptase polymerase chain reaction to look at the expression of the mRNA which uses the enzyme reverse transcriptase and the same technique which we just talked about polymerase chain reaction and Agarose gel electrophoresis these are all common techniques being used for doing this process.

So your interest is now to look at the expression at the mRNA level but to do that you are actually making the complementary DNA cDNA okay.

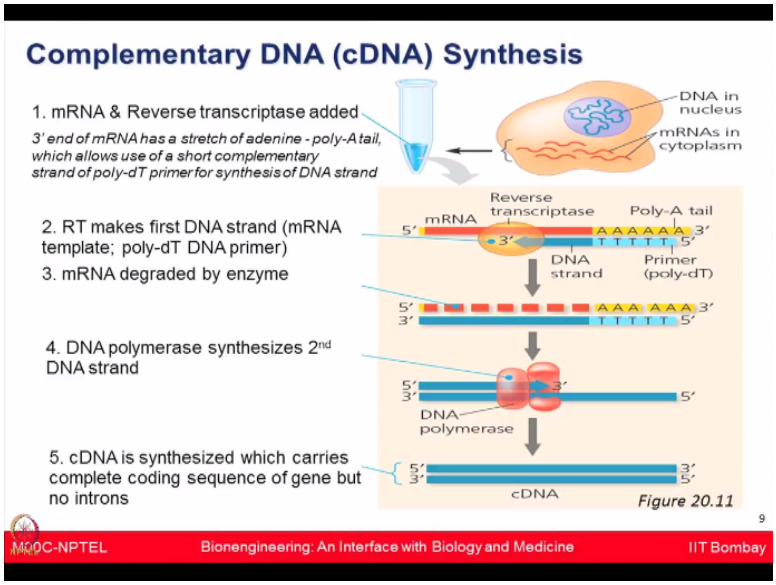
(Refer Slide Time: 09:47)



This is little complex but let us kind of go through it. So you have mRNA population in this cell. That mRNA can be converted into the complementary DNAs cDNA and I will talk to you about how that can be done. Now from this cDNA you can do the PCR amplification for given gene of interest which you want to see amplification happening.

And then those you can do further resolve on the gel to see that how much specific that band is, it corresponds to the right gene or not alright. So then this involves 3 major steps of first is doing the cDNA synthesis, second is PCR amplification and third is the gel electrophoresis right.

(Refer Slide Time: 10:29)



So the first part is most crucial here, let us look at that in some more detail. So in the cell we have all the mRNA and now we have to add certain reagents, certain components which can

convert that mRNA to the cDNA form right. So the very first thing is you can add the reverse transcriptase. Now this reverse transcriptase is ideally going to be involved in converting this mRNA into the cDNA right.

Just imagine in any given mRNA sequence, the characteristic feature is that you have a poly A tail. It means you have multiple of A bases in a you know continuous stretch. So this is what you can see here are the poly A tail. So you have many A's together. Now this property is being used to identify the mRNA's because now if you make a primer and if that primer has OligodT sequence, it means it has multiple of T's in a row.

So it is going to go and bind to that mRNA, so therefore you are specifically targeting the mRNA in this case. You have also added the reverse transcriptase enzyme and the poly-dT primers. So that is the first thing, then now the reverse transcriptase is going to start making the first DNA strand. So the red one was your RNA and now you are making a DNA from it which is blue in color okay.

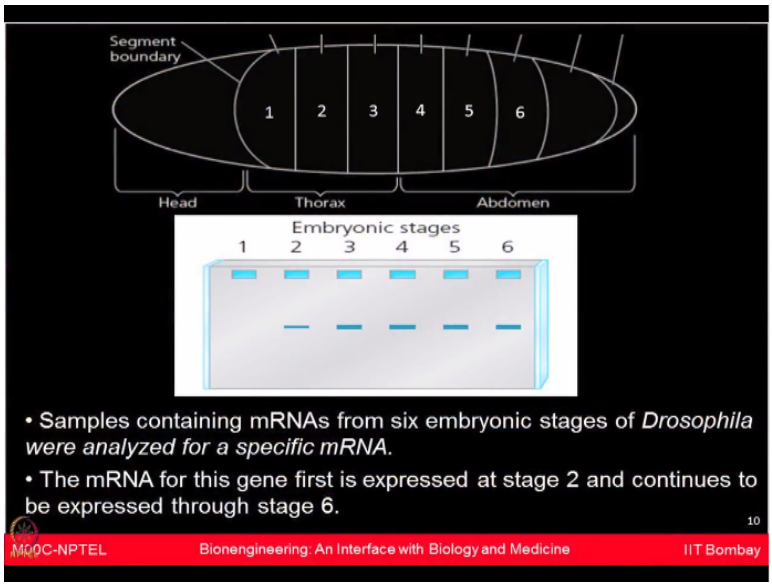
So now slowly because of this enzyme and the primer which you have already provided, now you can start seeing the synthesis happening at the DNA level and now you are adding certain enzyme like RNase H which could degrade this particular mRNA. So now because you do not want a population of both RNA and DNA together, you finally just want the cDNA which is complementary DNA sequence.

So you want to degrade even the RNA sequence from it, so to do that you are adding some enzymes which can degrade mRNA and then you are adding DNA polymerase which is helping to synthesize the second strand of the DNA. So first strand was already synthesized here, you degraded mRNA and now because of DNA polymerase activity the second strand is being synthesized.

So based on the second DNA strand synthesis which will happen with DNA polymerase now you can see the full of the cDNA can be made. So in this manner from the starting from mRNA you can convert that whole mRNA into the cDNA. Now the major difference between cDNA and the normal DNA will be a normal DNA will have what and the cDNA which will not have something.

Somebody rightly mentioned introns. So you have both exons and introns in the normal DNA when you have and now these introns being removed and now all these exons are being stitched together okay.

(Refer Slide Time: 13:26)



This is what the major difference you will realize between complementary DNA and the general DNA. So here if you see here in this Drosophila case, we were looking at various gene expression which is happening. So last time we looked at the mRNA profile so in turn from the intergenic regions they also have some genes but these are not going to be formed for the overall protein production.

So you want to get them cleaved off, so it is a normal process when it happens from the DNA to the RNA level you will have the introns chopped out as a part of the alternative splicing which happens and then only these exons are getting clubbed together which will form the functional RNA part. So in this case when we are looking at the RNA profile of Drosophila, if you remember we had various expression which we want to look for different genes.

In this case now let say we are looking at the expression level from the cDNA's right. So for those cDNA's now after doing the 3 steps of first was doing the cDNA synthesis, then polymerase chain reaction for that given gene and third is the Agarose gel electrophoresis. After doing all those 3 then you can see that what happens at the electrophoretic level for this gene of interest which you want to study.

Please do you know take the notes carefully because some of these steps you may have to write somewhere for a given experiment. So it means if the context is that you want to study only the expression of a given gene, it means you are looking at only the complementary DNA. To do that the very first step you have to go back to this particular working steps which we have talked.

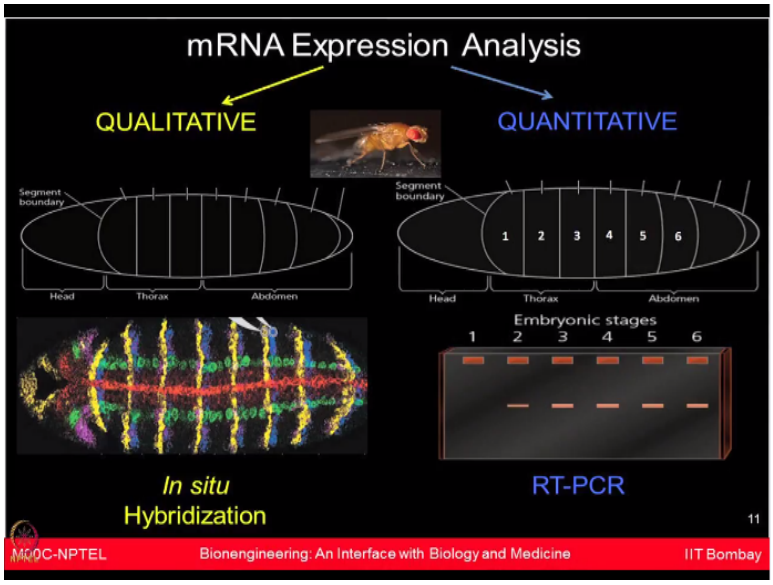
You have to go first on the mRNA convert and show how the cDNA's can be made and then for that given gene of interest you are designing the primers, you are amplifying that gene of interest using polymerase chain reaction at this level here and then you are amplifying that product you are running on the gel to find out is the right size for the given gene is being amplified.

So you are doing these whole steps in multiple kind of a stuff are involved in doing these kind of experiments. So please do write stepwise procedure if we ask you for a given experimental condition that how can you see that. So in this case I am just showing you a gel profile for a given gene expression but to achieve that particular gel profile there has been many steps which has been performed already.

So first you took mRNA population, converted cDNA, identified a gene of interest, amplified that using certain primers, did PCR. Now this PCR is known as RT-PCR reverse transcriptase polymerase chain reaction and then you are running on the Agarose gel electrophoresis to look for what are the bands which are being amplified here. So in this case one of the mRNA for the gene is not expressed in the very early stage but at the later part is actually being expressed.

And slowly you know initially the expression was very low but eventually you can see the expression was keep increasing. So these are the kind of thing which one could get an idea that for the developmental stage that which genes are present at which developmental stage.

(Refer Slide Time: 16:33)

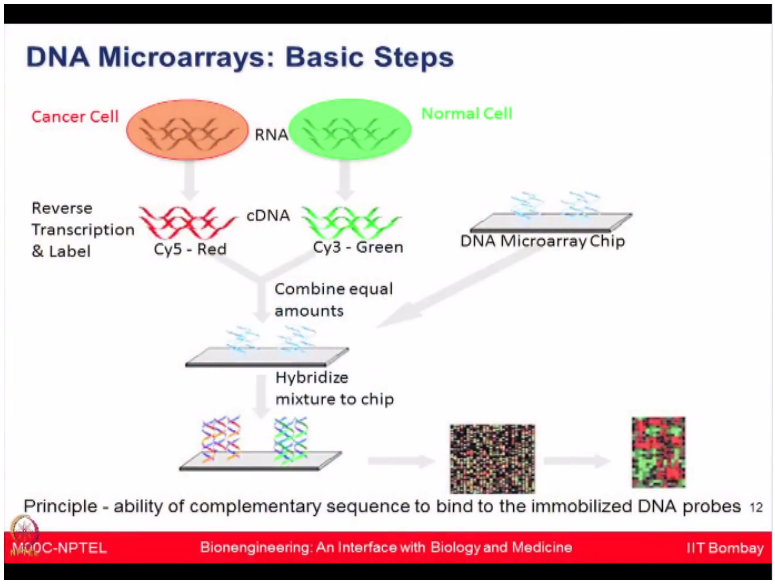


So broadly I talked to you about different ways of studying mRNA expression. One way we talked about qualitative methods especially in-situ hybridization, very beautiful images can be produced and you can see the RNA profiles here or you can also do the quantitative manner because you want to know exactly how much amount of that RNA which is expressed is present in the given cell.

And to do that you are using the reverse transcription PCR and latest addition of that is qPCR which is quantitative PCR or also known as real-time PCR where each of the gene enlargement or the amplification you can see the real-time manner. So that is much more latest addition which we are not talking right now but you know you can just think about both qualitative and quantitative manner of studying these genes and expression profiles.

One of the latest technique which is aiming to look at 1000s of gene expression simultaneously is microarrays.

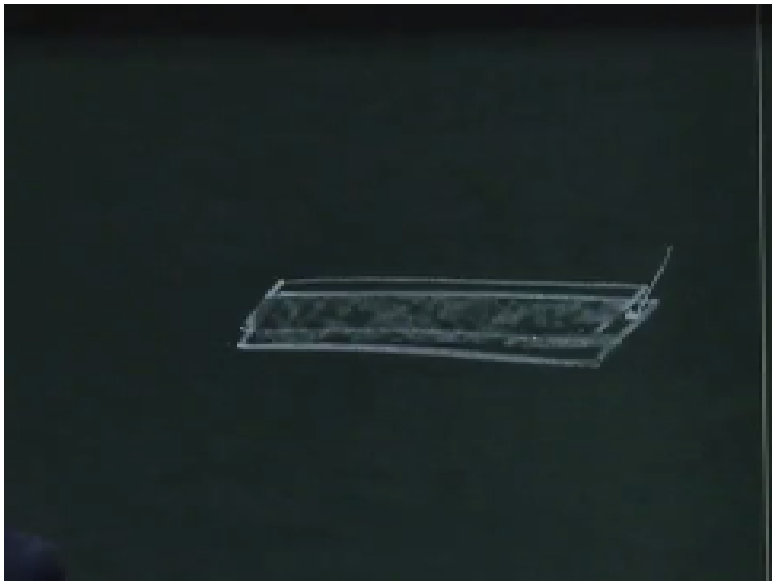
(Refer Slide Time: 17:32)



DNA microarray is one of the field which along with the other genome sequencing projects it was getting completed, people were able to synthesize those genes and then put on the chip. Just imagine you have a small glass slide like this size, if you have used any of these slides in microscope, if you have seen probably in our own experiment here though the small slights can have 1000s of small DNA oligos printed on them.

And now if you have very small amount of patient samples or clinical sample or any of the baluster sample to test out, you can use those and you can put that small amount of the samples on the whole chip.

(Refer Slide Time: 18:12)



So just imagine the chip can be so small glass slide like this but because you have 1000s of features printed here on the whole of this slide, you can put some coverslip on top of that, so

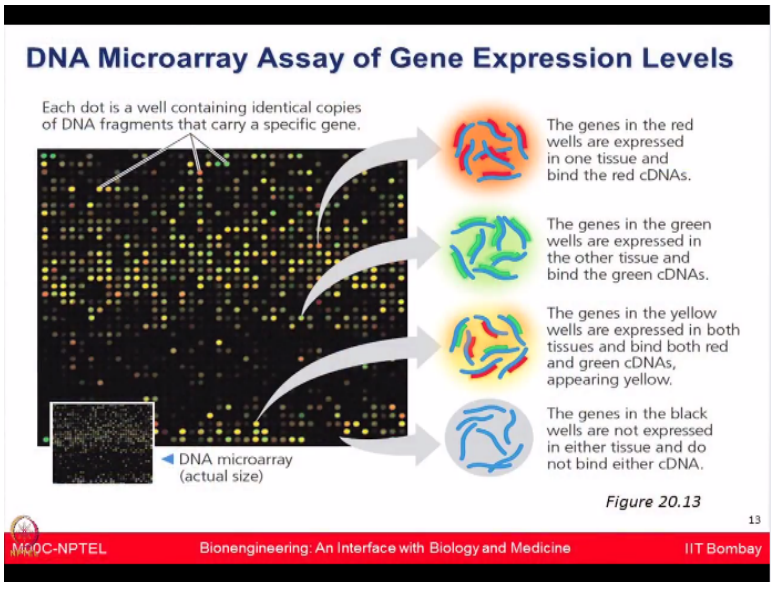
whatever biological sample you are adding here that biological sample will be uniformly flowed into all of the genes of interest. So at the same time you are actually able to study 1000s of gene and their expression level.

So whatever we were talking to you earlier was talking to you about one gene or one RNA at a time expression of that. Now these high-throughput technologies can give you ability to study 1000s in fact even millions of features at the same time. So in this method what people do let say we have a context when we want to compare normal cell versus cancer cell and we are looking at what are all the RNA present in the normal population, all the RNA present in the cancer cell population.

So all of this RNA you have converted to the cDNA complementary DNA which we have just done and while doing the cDNA in the last step you have added some tags, these tags at Cy3 or Cy5 colors and the result all of this particular normal cell will become let say green color and all of this particular RNA or the complementary DNA from it of the cancer cell will become red in color.

So now you have defined two colors already for two populations and now you are doing probing hybridization you are doing complementary probing of them on the gene chip here and wherever it goes and binds these RNAs or the complementary cDNA from it you are going to see the expression of those, how the expression of this particular gene is in the normal population versus cancer population.

(Refer Slide Time: 19:58)



So the principle here is based on the complementarity of the sequence and ability to bind with the immobilize DNA molecules. So what we see here we can see 4 different type of dots. If you look at this one microscopic slide here this pattern, after doing the experiment you can scan a different wavelength and you can generate these beautiful dot image colors and the rest of the things are just image processing part.

We had a lot of students from electrical engineering who had worked on the image processing problems of the microarrays lights. For them it is just simple image processing complex data which to be analyzed and of course for biologist what a useful significant information can come out of it can be so useful. So we have multiple dots, one is let say red in color, one is green in color, third context can be yellow in color and fourth is having the gray color or no color at all.

So when there is no color it means there was no expression of that gene at all. When yellow color is there, it means both the population from the cancer cell and normal cell they are both equally expressed and that is why neither green color you can see nor red color you can see and you can see yellow color which is mix of these two or you can have situations when you can see only RNA expression from the one of the condition which is shown in red in color.

So maybe some of the genes for the cancer cells are being highly expressed and that you see in red color or some of the genes are quite downregulated in the cancer cells and they are quite heavy or highly populated in the normal cells those are shown in the green color. So in this way just by looking at the expression of each of these genes and their color and ask this one could find out what can be the possible expressions of these genes.

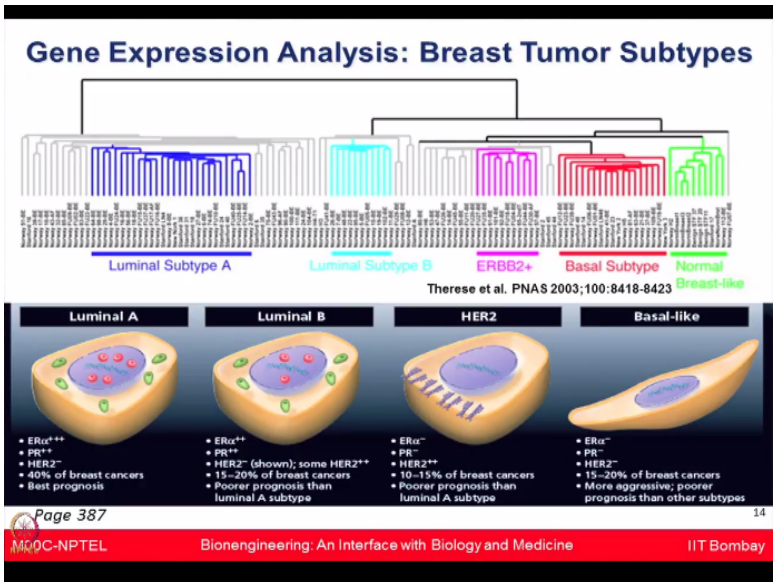
And you already know in this gene map you know you have simultaneously made a GAL file in the map which tells you that which are the location for each gene. So you can immediately tell that you know this particular gene could be that BCR-ABL gene, one of the mutant form of that and now this particular gene is showing very high expression in this particular type of cancer as compared to normal cell.

So in a very high-throughput way people can use these technologies. So this information of doing expression analysis has made tremendous change and progress in the way we see medical field today. So for example earlier people were thinking you know normal versus

cancer is one comparison but then people realize no each cancer is different type depending on their origin, pituitary cancer, prostate cancer or breast cancers are very different.

But then after doing the gene expression analysis for each cancer type people realize that each cancer itself is very heterogeneous.

(Refer Slide Time: 22:30)



So like breast cancer if you look at you know lot of women's who are all suffering from breast cancer, they have many different type of gene expression profile but there are certain patterns which are being formed across all these breast cancer women and those patterns could be grouped together into 4 major categories of luminal A, luminal B, HER2 and basal-like, it means all the women who are suffering from breast cancer they can have broadly one of these 4 patterns.

And these 4 patterns are based on the gene expression what we know for it and these are very different because if you are giving the same drug to all 4 type of group of women that may not be good thing for each one of them. So you need to give a different treatment here, you need to give different treatment here, a very different treatment here and different treatment here, so you need different drugs and different doses and different ways of treatment modalities for each subtype.

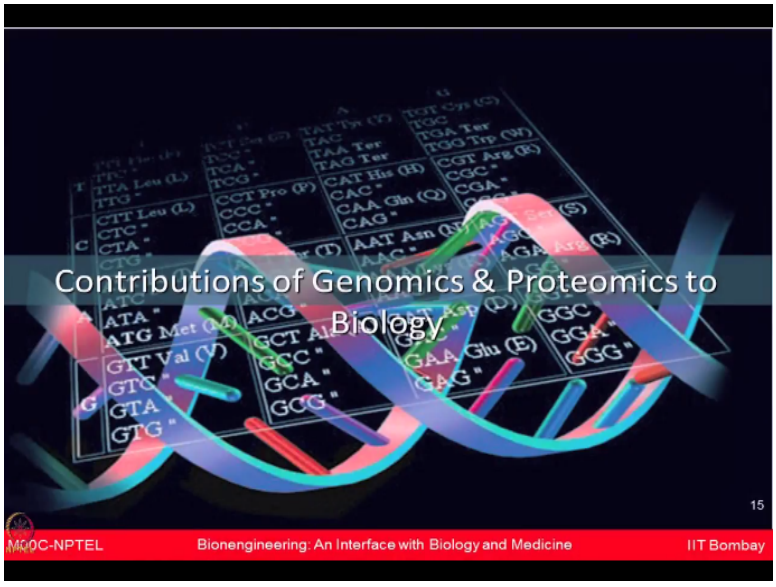
Being same cancer but still having so much difference in aggressiveness of the cancer or disease reoccurrence which may happen and all that we know now because of the analysis of gene expression. So for example in this case this particular luminal A type has very high

estrogen receptor expression of that genes. This is HER2 phenotype which is having very high level of HER2 genes.

And in this case we have again both high level of the progesterone receptors which is negative in the HER2. So just by looking at various genes and giving different kind of expression numbers, double plus or triple plus etc here you can get an idea that you know certain genes are having differential expression across different subtypes and this particular type of phenotype sometime these women's are known as triple negative when they do not have receptors for either estrogen receptors, progesterone receptors or HER2.

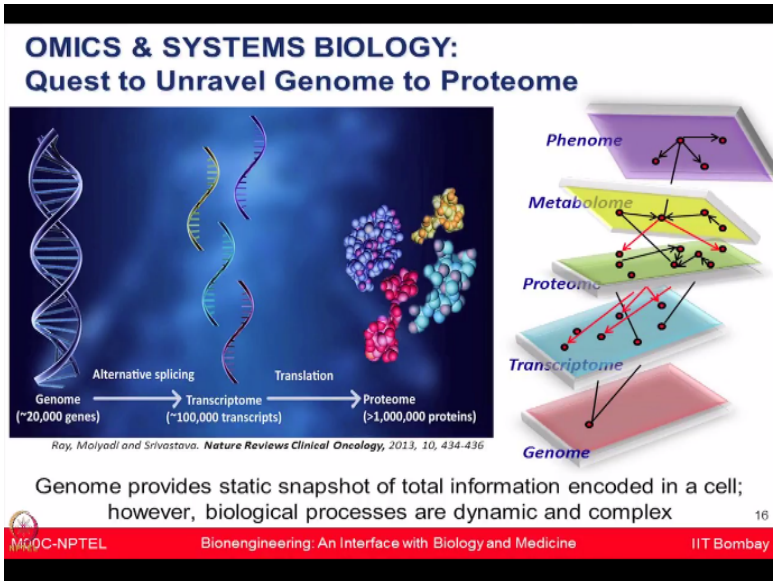
And then giving them treatment because much more tedious and they are much more aggressive in many context. So knowing these details by doing these kind of gene expression analysis has made tremendous change in the medical field and how we now do the patient treatments. So all these theoretical knowledge which we have been talking has actually started making a good impact in the medical field today.

(Refer Slide Time: 24:29)



So therefore although this is not kind of I am going to talk in much more detail but understanding about all the genes of a given system or knowing all the proteins of a given system as a part of genomics or proteomics has been very crucial and fundamental to biology field.

(Refer Slide Time: 24:42)



And it has because if you just think about the central dogma from the genome to the entire transcripts being form of the transcriptome to all the proteins being form all the proteome you have all the you know biological processes being governed here and if you can study all these cell population, all of their RNA and all the proteins which are being expressed or being changed then probably you are better informed to know that what is happening in that system as compared to you are only looking at one or two gene as a protein at a time.

So this particular field has really benefitted quite a bit also from the computational analysis where now our aim is that we have information for all the genes for a given system. We also know which are all transcripts being produced, how many proteins are expressed from them, what are all metabolites being present and how a phenotype of that particular organism or system looks like.

Can we combine all this information because we are talking about same system right? Let us think about human, so we are talking about same human which are all genes present, all proteins present, all metabolites present and how our behavior and phenotype gets changed in different type of situation. So can we study all of these properties numerically in a format which is known as system biology.

It means we are studying a system, we are deriving models from each layer and trying to stitch it together and bring some very new and novel information for that system and that is known as systems biology which is very crucial in this entire field.

(Refer Slide Time: 26:11)

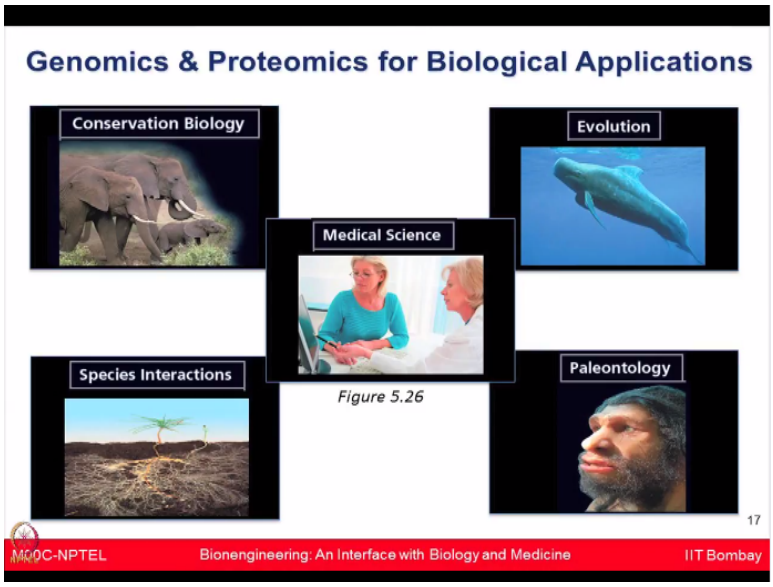


Figure 5.26

So these are some of the examples from your book where you know lot of revolution has happened because of our advancement in understanding genome and proteome and these are apparent in every single field whether you think about even conservation biology, look at evolutionary studies, species interactions, paleontology and of course I have been talking much more in the medical science context.

All of these fields have benefited heavily for knowing that you know what can happen at all the gene sequence level and all the protein levels. Today, we learnt about some very essential molecular biology tools that can help us to study gene expression. Some of these technologies are so generic that it could be employed for any biological applications, whatever questions you might want to address and you want to look at the perturbations of genes they are going up or down regulation in response to given stress, in response to any disease condition.

All of those things are possible by employing these kind of biotechnology tools. We also learnt about the significance of learning the differential gene expression and how it could aid in the patient diagnosis. Furthermore, we now know that these tools are of paramount importance for the clinical decision making especially you know for the drug delivery and many therapeutic interventions.

We have also learnt about the system biology where a lot of computational tools can be used to solve key biological problems. Thank you very much.

(Refer Slide Time: 27:41)

Summary

- DNA tools and biotechnology have revolutionized diagnostics and therapeutic interventions
- We can now easily sequence thousands of gene in cost effective manner
- Integration of technology with biology has taken the field forward.

(Refer Slide Time: 27:45)

References

• Campbell Biology - Reece, Urry, Cain, Wasserman, Minorsky, Jackson 10th Edition, Cummings