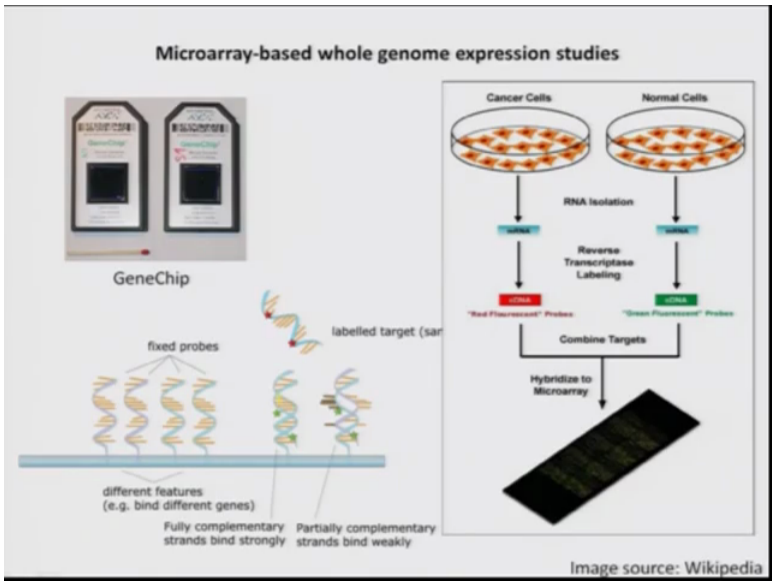
Functional Genomics
Professor S Ganesh
Department of Biological Sciences & Bioengineering
Indian Institute of Technology Kanpur
Lecture No 08
Transcriptomics Part 2

Welcome back to the course on Functional Genomics.

(Refer Slide Time: 0:23)



So let us look into this scenario for example you want to look at the question is that you have a cancer tumor tissue and you have the same normal tissue from where the tumor has grown. Now we want to look at the tumor condition what are the gene that are upregulated whose expression level has gone up, which could possibly is involved in some tumor formation. And likewise what are the gene which are normally expressed in the normal cells but now shutdown in the tumor cell. That means that cell gene is required for the normal function right.

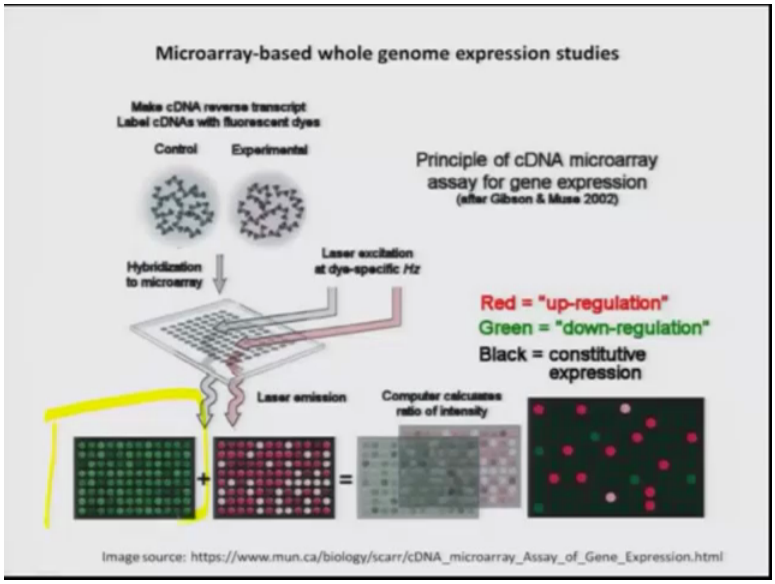
So what you do, so you have these two samples we extract the MRNA from each one of them and then you make what is called as cDNA that is something that we have discussed so you basically you make a complimentary copied DNA. And when you make a cDNA, what you do is you tag the DNA being made with certain florescent mites, so let us say I will use I am going to use a florescent mite that gives red florescence when I excite with certain wavelength for the cDNA that derived from cancer cell or I can use for example for the normal cell I am going to

use a florescent mite that gives you green if I excite with a particular laser or a wavelength of light.

Now what I do is now I am going to take equal ratio of these two cDNA so for example 2 microgram each I am going to mix and then this is my gene chip something that shown here, now I am going to hybridize, am going to allow this cDNA because now I got rid of RNA from here, now you only have one single stranded cDNA which is complementary to the sequence that are spotted there. If I have ton copies of a given genes spotted on a chip and if the expression of a gene is same in both cancer and normal cell then going by the probability 5 copies of the gene will be bound by the cDNA derived from the tumor cell and 5 copies of the target sequence will be bound by the cDNA derived from your cancer cell that is way it is.

But if it is the equilibrium shifted you have more of that transcript plus in the cancer cells you are going to have more of cDNA that is going to go and bind to the spot. So by looking at the relative ratio of these florescence I will be able to tell whether the expression level is altered or not altered, that is what is shown here.

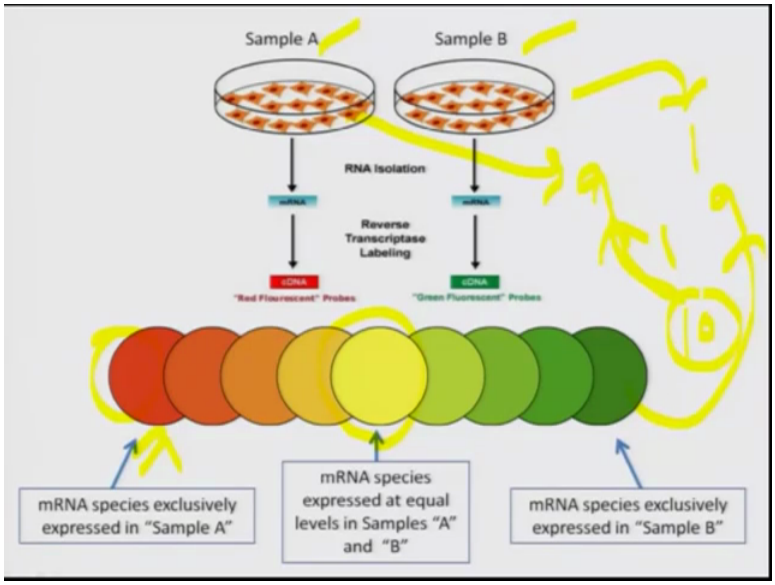
(Refer Slide Time: 2:57)



For example this is your array you have spots there so what I am doing is that once these are arrays are hybridized then I want to quantify the florescence by you know in exposing that chip to a laser source, excite that in a florescence mites therefore it emits lights and I am going to

measure that. So what I have done is first I have measured the green then I have measured for the red I am going to look into the relative expression of the red and green which will tell me whether as compared to the normal cell in the cancer cell whether the expression level has gone up or gone down or not expressed and this is we are going to do.

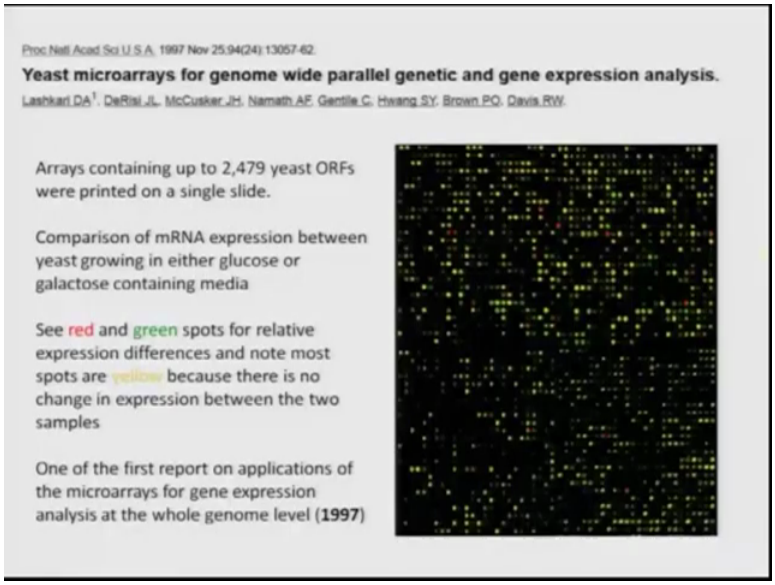
(Refer Slide Time: 3:40)



So let us see that this is what it is. So if you know the expression level remains same pretty even between the two source Sample A and Sample B then I would get something like green yellow because this is a mixture of red and green right. Whereas the expression level in sample A for a given genes very high I am going to get red florescence because the green florescence coming from sample B because it is a competitive hybridization. Say suppose for every one transcript for a given gene from sample B you have 9 transcript from Sample A and I have 10 targets here. So 9 of them with be bond with the cDNA derived from sample A, so we are going to have more of red color that is what it is.

On the other hand if its reverse 9 here and 1 here I am going to get more of green. So by looking at the intensity of this 2 fluorochrome the light emitting mite I am able to interrupt it whether the expression has gone down gone up remained the same right so this is the way we are able to do that. But you are not doing it just for one gene, in 1 go you are able to do it for thousands and thousands of the gene with the little amount of RNA.

(Refer Slide Time: 5:05)



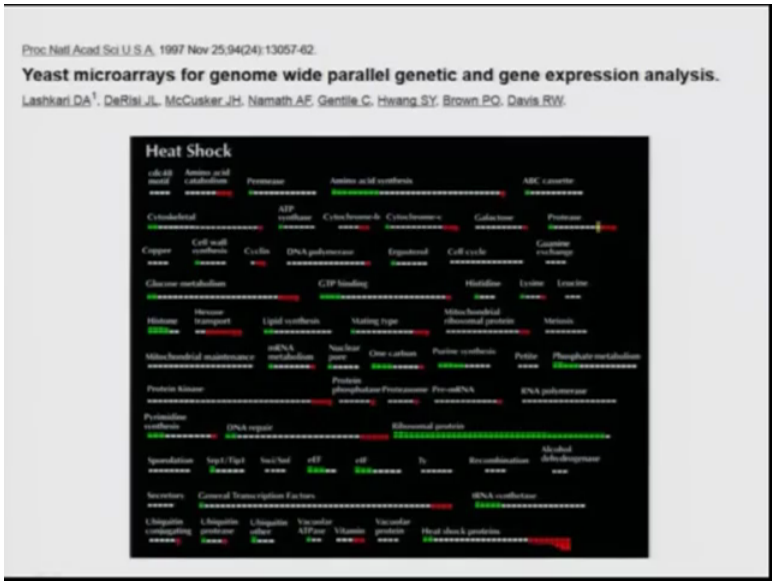
So that is the power of the microarray and it has really changed the way we look at gene expressions and this is one of the landmark of the (())(5:04) perhaps the first paper looked at the expression in a genome wide basis. This came in 1997 where they looked at the yeast you know the popular model and its expression analysis which is called as Genome-wide. You can see that now this is being more often used but that is the first time they have used, so Genome-wide Expression Studies.

So what they have looked at, so they have looked at you know relative expression of genes if we know on of yeast that were grown in 2 different medium. 1 is glucose and other one is galactose okay. And they looked at what are the genes that are turned ON if you grow the yeast in glucose and what are the genes that are turned ON if you grow them on Galactose. So they looked at how the profile expression profile of all the genes changes simply by modulating the culture condition and how many genes they have looked at close to 2500 genes in 1997, it is a huge you know shift from the way people were looking at it.

And then it is not only expression they are looking at relative expression that is what the power of microarray. So it is not just saying expressed or not expressed, whether its expression level is higher as compared to this condition or lower as compared to this condition and you can see that there are dots that are completely green, there are dots that are completely red but majority of yellow suggesting that you know majority genes pretty much did not change the expression but

some of them shutdown, some of them activated and so ON. So that's the first report on the application of microarrays for gene expression at the world genome level came in 1997.

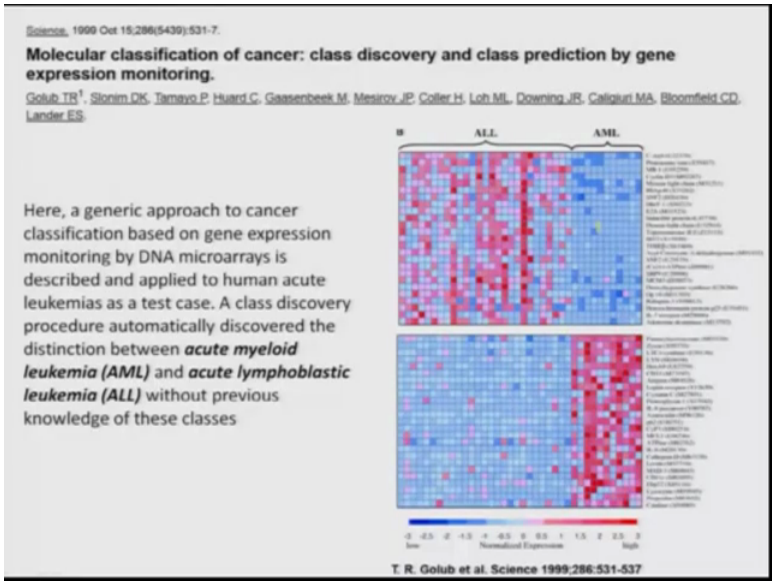
(Refer Slide Time: 7:01)



Since then you know there are large number of papers. But what is interesting here is that this kind of analysis gives you is not you have not really looked at what gene it is. You are smart in every damn gene that you have it and you are going to look at whether the RNA is made or not made. Now having found a difference now that they are over expressed or under expressed now you can go and look at what is the function of the gene, what is the pathway and then how that may make a sense, because there are full of data, you have to analyze this.

So this is what they are showing here, for example there are protein if you give a Heat Shock for example whether (())(7:33) metabolism is altered, whether that are gone down gone up and so on, so this is the kind of post analysis which gives you much larger picture which can help you to come up with a new hypothesis test, you know the function. So that is another paper in 1999.

(Refer Slide Time: 7:53)



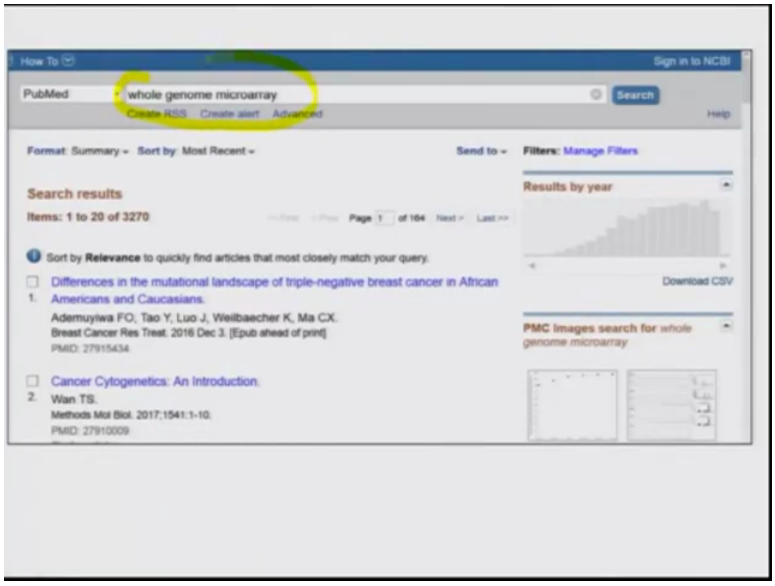
Again the landmark paper they have looked at cancer because one of the biggest challenge in cancer is that you have different grades of cancer subtypes of cancer. So your treatment basically depends on if you are able to diagnose with subtype of cancer that one person is suffering from and this is always a challenge as to how do you classify. So most often people use histology looking at but these are much changes that you see over there is very late by then cancer is really progressed to much severe level. But if (())(8:22) early it is better, so to type that can you look at the expression profile, can you look what are the genes that are expressed, can you type them that is type 1 or type 2 and vice versa and then one can come up with certain therapeutic modules.

And here what they have done is they are going to take a approach for cancer classification based on gene expression monitoring DNA microarrays use this micro array to you know look at the gene expression and then they what they have looked at is that leukemia is the model and they have looked at 2 types acute myeloid leukemia, acute lymphoblastic leukemia these are subtypes and they are able to show for example there are certain genes that are present or expressed or you know in a very distinct way as compared to AML you know ALL is the acute lymphoblastic leukemia AML is acute myeloid leukemia, so they are able to type now.

So this way they are able to tell that this tumor belongs to which subtype you know if there are drugs that would be more affective in AML this person can strait away be given AML without

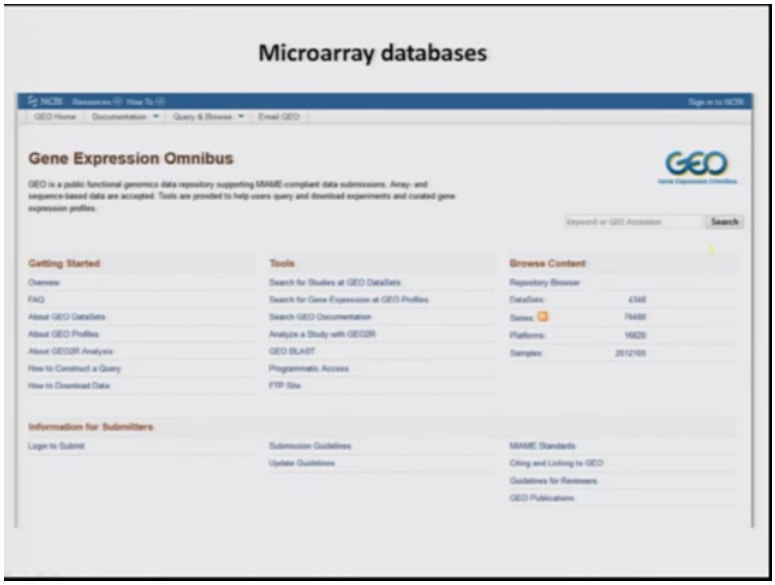
really waiting for further confirmation. So this is again a landmark paper which really helped us to classify the tumor and go on.

(Refer Slide Time: 9:54)



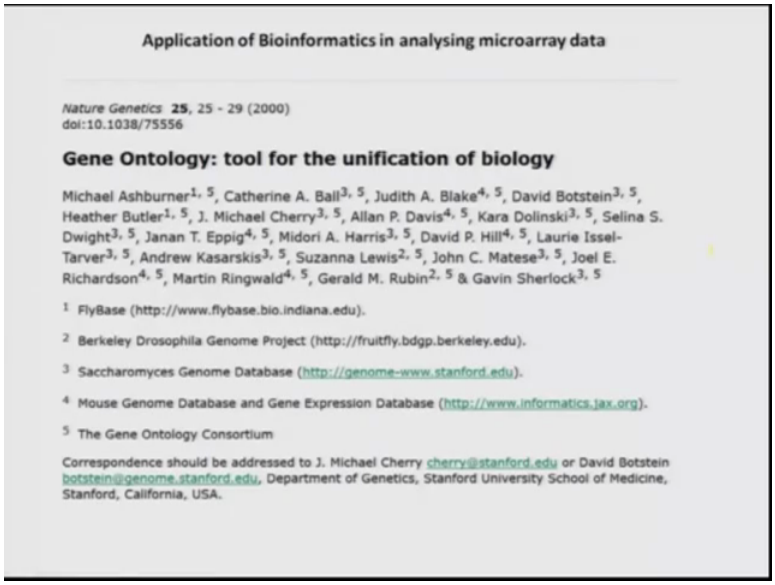
And if you look at the PubMed again if you give the query as Whole Genome Microarray and you can see that then how many studies are there right, and it is only growing with time and with time it is going to grow up. And that's the application and this also bring in a challenge. The challenge is, what does it mean, so you are spotting all the gene that are predicted that identified the genome for many you don't know what is the function. So if you get some change as the microarray suggest that expression level gone up or gone down. What really you know it would mean.

(Refer Slide Time: 10:36)



So this really requires you know not a manual analysis, it requires computational tools and that is where you know researchers have found a need that there are all these data that comes out of the microarray, the fold difference in the expression level and so on should be deposited and in a common repository, because even the tools to analyze them are evolving so if the data is present then one can evolve tools to analyze that. So what is shown on the screen is one of the data bases called Geo Omnibus which is nothing but Gene Expression Omnibus and this is a that in a database people deposit their expression analysis data there and anybody can use it.

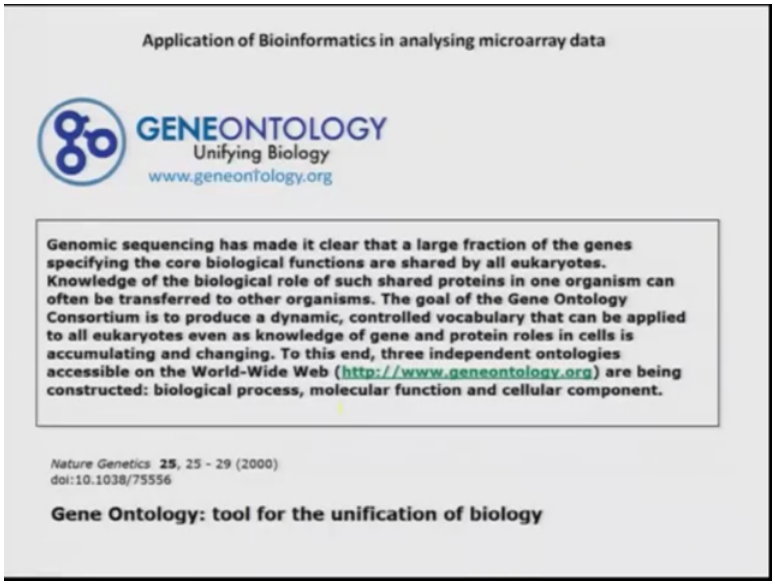
(Refer Slide Time: 11:19)



So that led to a new branch of Bioinformatic is called as Gene Ontology meaning what is it mean. So say suppose I am getting a gene whose expression level is altered. What does it really mean in terms of the biology, because (())(11:25) living at biology as to how that gene has altered the function for example cancer progression or its you know whether a drug is able to rather the tumor is able to respond to the drug or not. So all these people who are involved in such kind of you know high throughput genome wide microarray analysis they came together and and sort of came up with an idea of that that you have to have some tools to classify you know the information that you get out of microarray and that led to a bioinformatics powerful bioinformatics tool called Gene Ontology.

And that is now it is being widely used not only for human you can see that whether it is for mouse or yeast or you know drosophila all these people came together and I know came up with a uniform way of classifying genes based on function. And this is a dynamic function because the gene being studies today and you have you know that this is involved in this function the database is updated. So as the results come out this database is being a constantly updated and this is called as the Gene Ontology Consortium or Gene Ontology.

(Refer Slide Time: 12:38)



So that is the website that all of you can go on and look at. And when they formed this consortium this is a mandate, it says I read it out. Genomic Sequencing has made it clear that a large factor of the genes specifying the core biological functions are shared by eukaryotes whether it is a DNA Palomares Histone Protein they all do the same function where they are looking at yeast on the human. Knowledge of the biological role of such shared proteins in one organism can often be transferred to other organism because if you are looking at a developmental process you will never be able to study that in human.

But you can study that in zebra fish or fly and that information should help us to understand that genes present in the humans. The goal of the gene ontology consortium is to produce a dynamic meaning dynamic it has to be updated every day because the research is going on, control vocabulary should not be something that everyone uses a different terminology it should be the same if it defines the same function therefore easy to understand that can be applied to all eukaryotes even the knowledge of the gene protein roles in cell is accumulating and changing, because you have to have some unifying code.

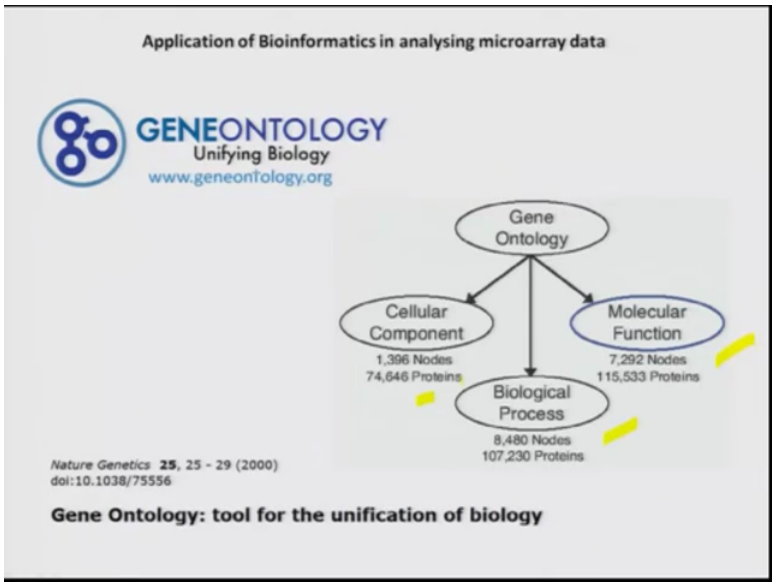
To this end, 3 independent ontologies accessible on the world wide web that is geneontology.org are being constructed that is when they put this proposal now it is pretty much you know is being used. These ontologies are grouped into 3 distinct domain, 1 is biological process 2 is molecular function, cellular component. For example I may be looking at a protein which is a transcription

factor. So the function of the protein is DNA binding where it should express, it should be present in the compartment which is nucleus.

So this protein give 2 information number 1 it is present in the nucleus 2 it is DNA binding in its activity. What is the third function, it may be involved in skeletal development so that is the process right. So this would evolve as people are studying it will evolve. So they are keep on updating this database. So when I do a microarray for a given tissue or whatever process, I found a gene, this is the gene whose expression level is very high then immediately I want to know what does it mean, you know level gone up what does it mean.

So this is the protein that is present in the nucleus therefore it should be high if a DNA binding protein and therefore it should because activating or inactivating a transcription. And what process it may affect for example skeletal development right, so this is how you look at it.

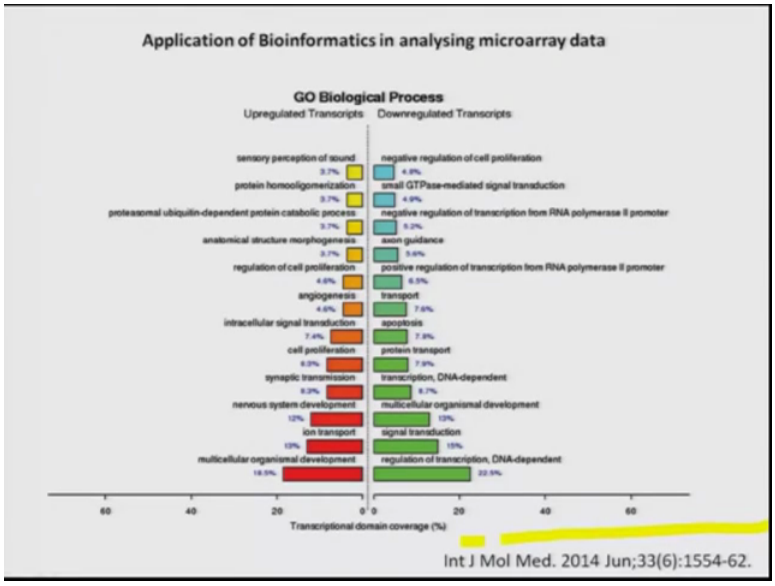
(Refer Slide Time: 15:23)



So this is what it is shown and each you know you know theme as you know sub notes you know because that within molecular function it could have enzymes you know you could have many things. Biological process likewise we will see some of them. Compartment it could be nucleus, within that it could be membrane and others so on. So they have classified it so as the information evolves it gets updated you may have (())(15:39) to go to this ontology website and have a look at it. I am going to give you examples of 3 case studies. We are not going into why

this study was done if you are interested you can go and look at this particular paper. What I am showing is how you know they have analyzed the data.

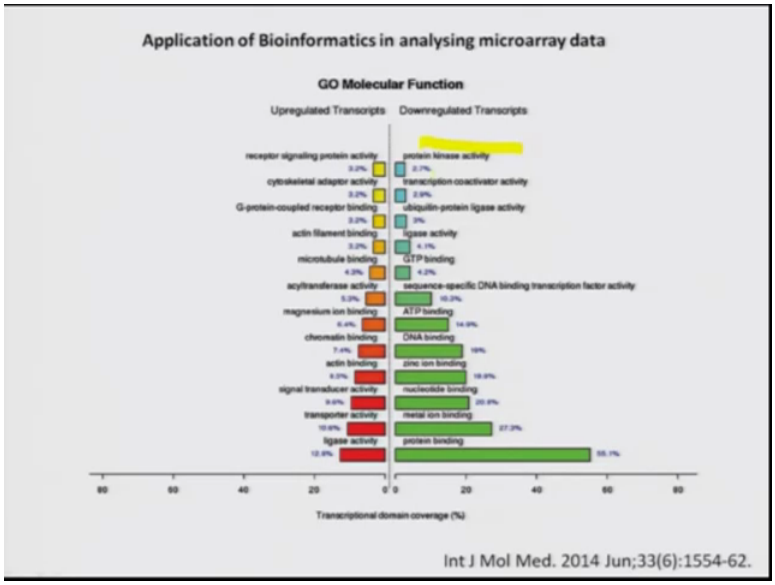
(Refer Slide Time: 15:59)



Let us look at they have done a microarray you know between 2 different samples looked at. And there are certain genes whose expression level came down whose expression level went up and they are looking at what is called biological process right. What is biological process? Sensory perception 3.7 percent of the genes whose expression level you know came down or involved in sensory perception like touch smell whatever. Protein oligomerization protism living in different protein catabolic process, these genes whose protein involved in proteolytic process regulation of cell proliferation controlling cell division and geo genesis meaning blood vessel formation, cell proliferation, synaptic proliferation.

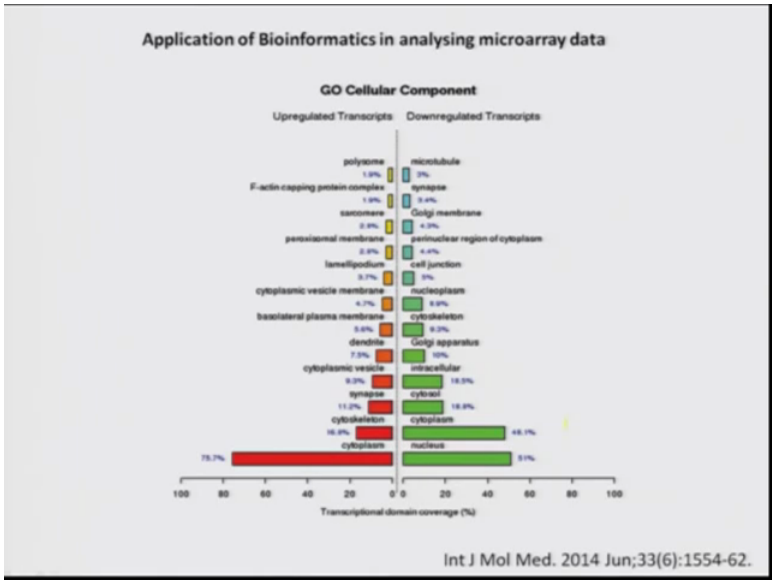
You can see that how we can say that okay is it that this is very interesting you know almost close to is 20 percent of the genes whose expression levels are up regulated or involved in multi-cellular organismal development that means sub growth happening right. You can go back now look at the paper what did they do, why possibly over expression of the gene led to that process, it will be interesting and you see that there are down related transcript regulation transcription DNA dependent and so on. So this is how you can classify.

(Refer Slide Time: 17:27)



Molecular function again, this is a second sub-theme then you see that there are some of kinaes activities some are transcription factors ligase activity, GTB binding, DNA binding and so on. What is the specific function of that? Protein but the pathway could be different but we are looking at a specific function or you can go at the level of cell as to where it is located.

(Refer Slide Time: 17:43)

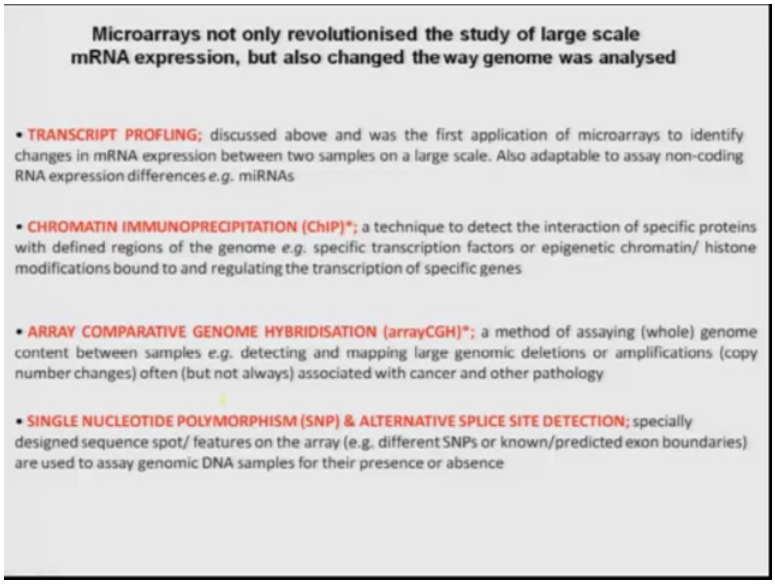


For example is that present in microtubule, nucleoplasm, golgi apparatus, cytoplasm, nucleus and you have information. As it is evolving if somebody has done in yeast and you have a protein

which is a homolog meaning either the protein that represent the yeast gene in the human it is likely that it does function. So it is indicative, it is not saying that that is absolutely that is the function but it is indicative. That's how you go and further modulate and test it in the system.

So anybody is working on any system now everything is put on website linked to that gene and protein and you are able to get without really getting into the paper if you are looking at you know thousands and thousands papers that are there whose results are sort of linked to all these genes, you do a microarray you get all these data. So it is a powerful tool made possible by using bioinformatics approach and that is being constantly updated by looking into the literature. Again there are some bit of human element involved, it is not everything is automated but certainly you know one group of people is doing it for the benefit of the large number of scientific community right.

(Refer Slide Time: 19:05)

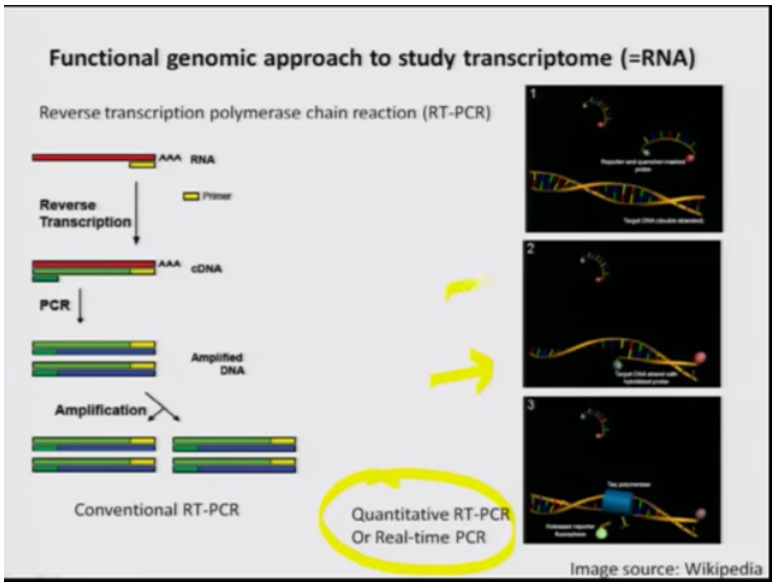


So the microarrays not only revolutionized the study of large scale MRN expression that is what we have discussed but also has changed the way the genome was analyzed right there are you know you know changes at the genome level. For example you know there could be repeat regions, repeat numbers and increase or it could have you know a small segment of genome is duplicated 3 copies and so on. So one word you have looked at is the transcript profiling we have so far looked at MRNA but now you have such kind of arrays for non-coding RNAs micro RNA and so on.

The other thing is to look at the changes in the DNA that makes a gene to be active or inactive. This is called as Chromatin Immunoprecipitation something which we will discuss little later. And there are also arrays to understand what is called as Comparative Genome Hybridization. For example whether a region you know of genome whether that has got duplicated.

The copy number increased or last, that can be looked at. It is involved in large number of pathology including cancer, we will study little later. And of course sequence variation you know what is called as Single Nucleotide Variations large scale. And also Alternative Spice Site as we discussed that each gene has multiple splice form. Now you have arrays that can detect all these things.

(Refer Slide Time: 20:25)



So before you end this particular lecture we will look at one last approach. Again we come back to this real reverse transcription polymerase chain reaction. And in this what we are going to do is microarrays is indicative of difference in the expression but if you really want to quantify people have to come and look at individual genes to quantify them. So this PCR we already discussed, you are going to amplify a segment of cDNA but that does not give you the you know the difference you want to really quantify the difference of the expression level between two difference samples for same MRNA.

Now a days we use a technique called as Quantitative PCR or Real Time PCR. Here it is the same principle like a PCR one of the ways we can detect is to look for the number of duplex DNA being made you know because more the number of duplex DNA in a given number of cycle that means you have had more copies to be in with. So if you are comparing 2 different samples then you will be able to tell whether you had more transcript as compared to other.

The idea is that you have what is called as one of the way which you study is you have the proofs which are shown here the proofs bind to your single standard DNA and they have certain fluorochromes and the fluorochromes are linked to that proof because is it you know you have oligonucleotide this fluorochromes are there normally when they are together they would not flures but when even if the bond formed if the DNA (())(21:56) is come and this (())(21:58) is done in that process they (())(22:00) sort of break them then you get the fluorescence that would tell you how many of them are bound, how many of them converted, that would give you a relative estimate as to how many copies you started with.

With that you can you know identify, measure the relative difference in the transcription. This is a very accurate way by which you measure the transcript and that’s pretty much the end of this lecture where we looked into how people study a large scale analysis of the transcripts and you know in the next lecture we will move on to look at how we study the genome. Again we will go back to genome and its dynamics, significance how does it affect the transcription.