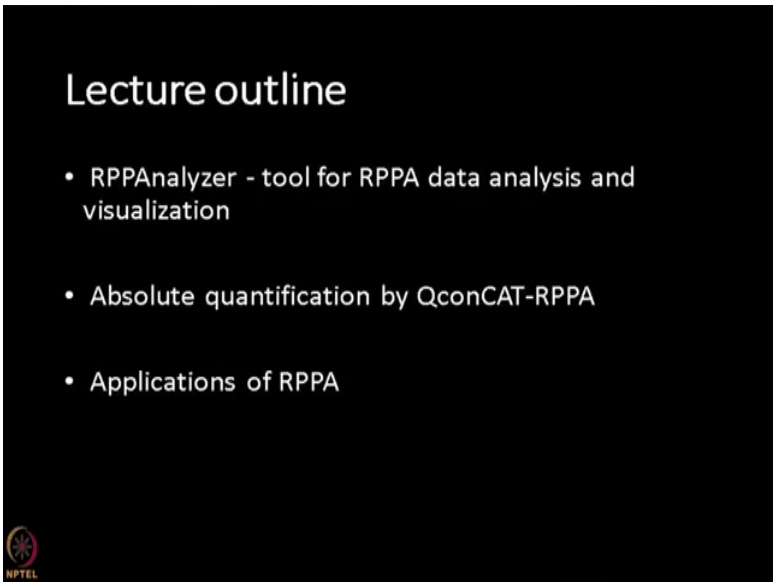
Interactomics: Basics and Applications
Prof. Sanjeeva Srivastava
Dr. Ramesh Ummanni
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay

Lecture – 27
Basics and Applications of Reverse Phase Protein Arrays – III

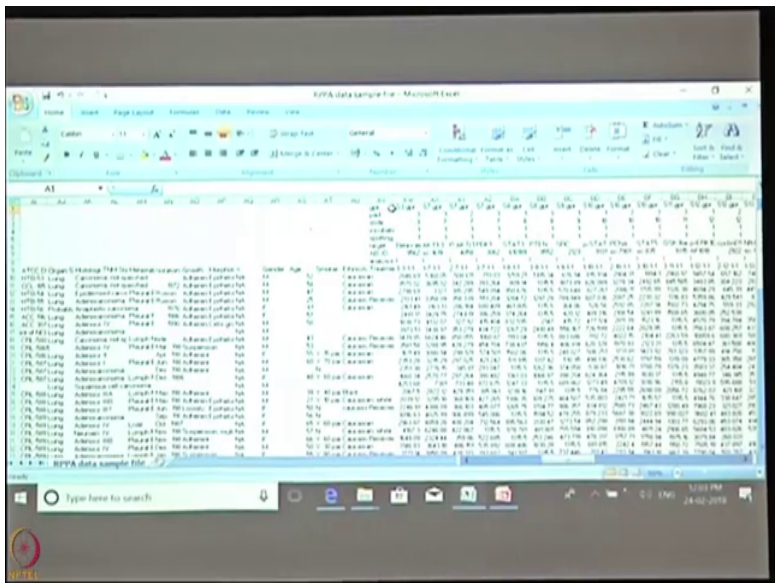
In continuation of doctor Ummanni's previous lectures where he talked to you about Basics of Reverse Phase Protein Arrays. We are continuing to discuss about this novel technology platform RPPA. In today's lecture where doctor Ummanni is going to mainly focus on RPPA analyzer; which is a tool for RPPA data analysis and visualization. So, let me welcome again DR. Ramesh Ummanni to continue his lecture and discussion on reverse phase protein arrays.

(Refer Slide Time: 00:59)



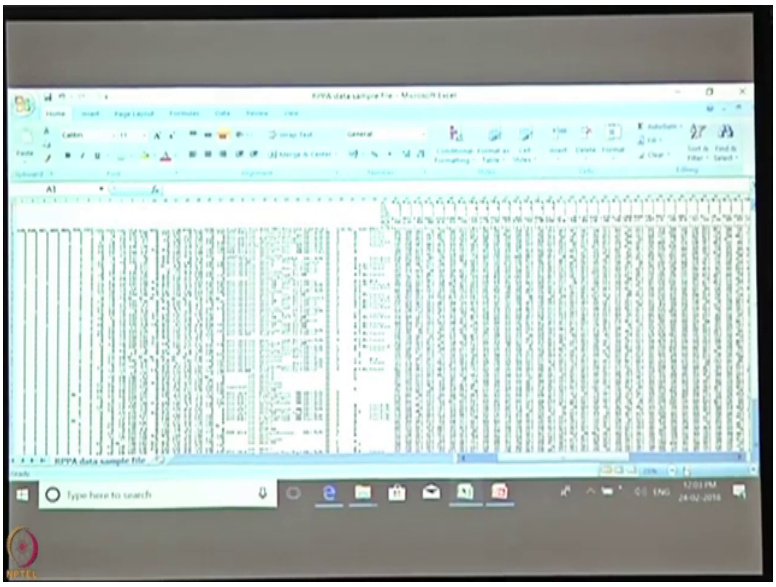
I will show you what kind of data we have generated. So, here let us say for example, these are the different samples, these are the different mutations in the cell lines. So, you can see that this is the data we could generate.

(Refer Slide Time: 01:20)



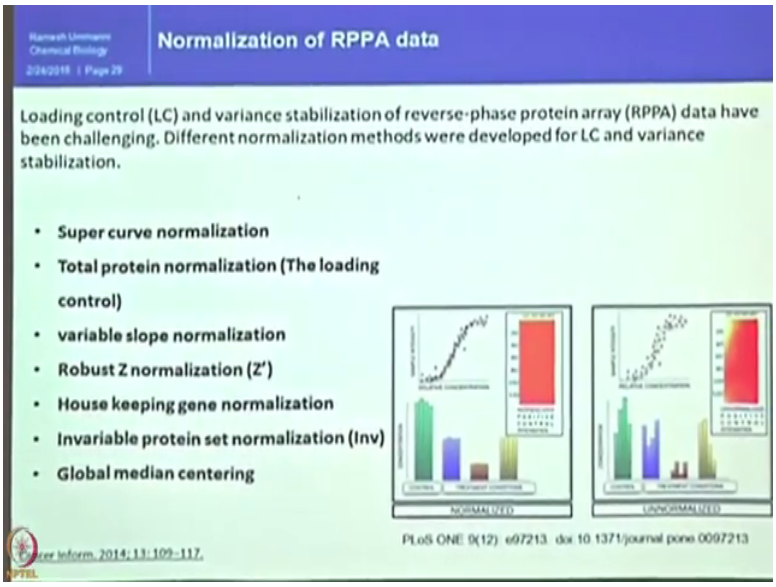
So, in this corner on the drug resistance can be incorporated here, drug resistance can be incorporated. So, this kind of numbers you will get. So, these are the target names.

(Refer Slide Time: 01:32)



So, your complex matrix will be like this. So, from this if you load this data into commercially available software. So, as per predesigned analysis part you will get the data output, but you cannot manage to get information what you want from your samples according to your experimental requirements.

(Refer Slide Time: 01:54)



So, when you have a such a high end data set then you need to worry that how experiment is correct, how much experiment is correct or not. So, one of the sensitive point that Sanjeeva touched in the morning very nicely and gently suggested you to follow this normalization methods.

Particularly in RPPA data, normalization is very very difficult because when you do the protein estimation or when you extract, all these steps cannot be implicated for 1000 samples in the same manner. It is very very difficult. If you do 10 samples, we will have ten different values in terms of protein concentration, in terms of method you followed and all.

So, to avoid that there are different normalization methods have been developed to apply for RPPA data. So, loading control and various stabilization of RPPA data have been challenging. So, different normalization methods were developed for LC; that is loading control and

variance stabilization. One among them is a super curve normalization, total protein normalization, the variable slope normalization and robust z score normalization methods.

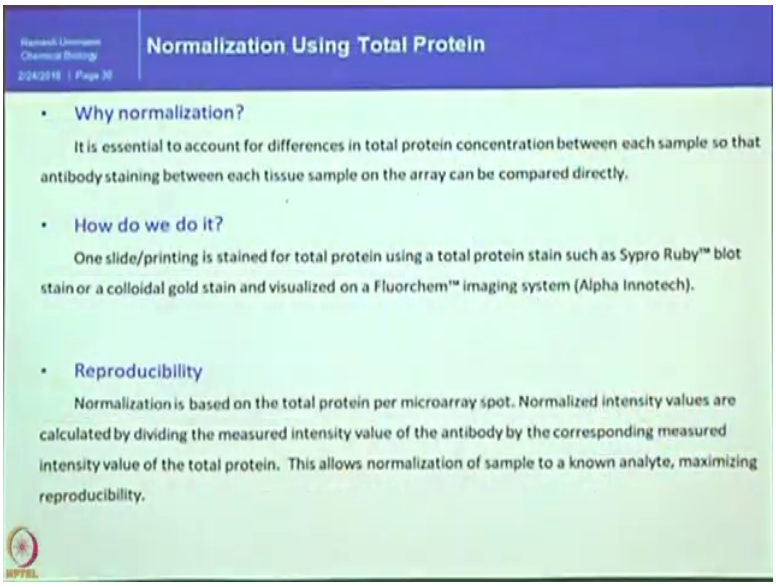
One of the method what we follow in western blot is housekeeping normalization. We all favorite of doing GAPDH more than the target proteins very often. And invariable protein set of normalizations; invariable protein set of in normalizations means; in all the 1000 samples we know that one particular protein concentration will not change.

So, that we will set as a landmark. I think if some of you might have done west sorry 2d gels in the proteomics approach, in the early in the beginning days. So, we used to set landmarks spots. So, in that way those proteins expected to not change across the sample set will be set as a landmarks spots.

Based on those concentrations we can normalize. How a normalize data look different from the un normalized data? You can clearly see that these bars, each bars represents one protein across the controls and treatment conditions then you can get a heat map like this. But after normalization, so there is no difference within the samples and nicely you can get a spread of the signals here.

So, if you do not apply any normalization method, you end up identifying the more differentially expressed proteins across the sample set you have analyzed.

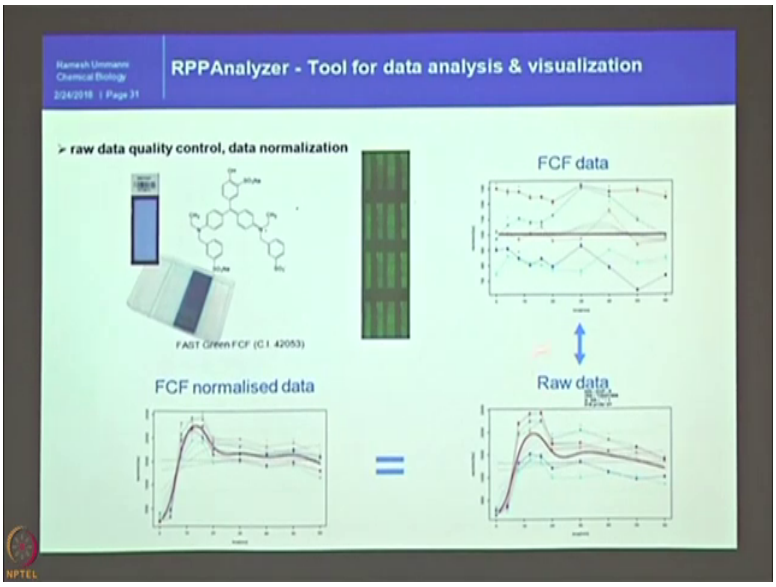
(Refer Slide Time: 04:11)



So, how can we do? As a among the methods previous slide explained, at least we use to follow total protein normalization method which is more easy to do and also reliable to do. So, it is how do we do that; always we print lot of slides. One slide after every 10 slides will be used for this normalization methods. Every 11th slide, 21st slide, 31st and 41 slides will be used as a normalizing slide, normalizing slides.

This will definitely give the reproducibility. This normalization is based on the total protein per microarray spot. How we do?

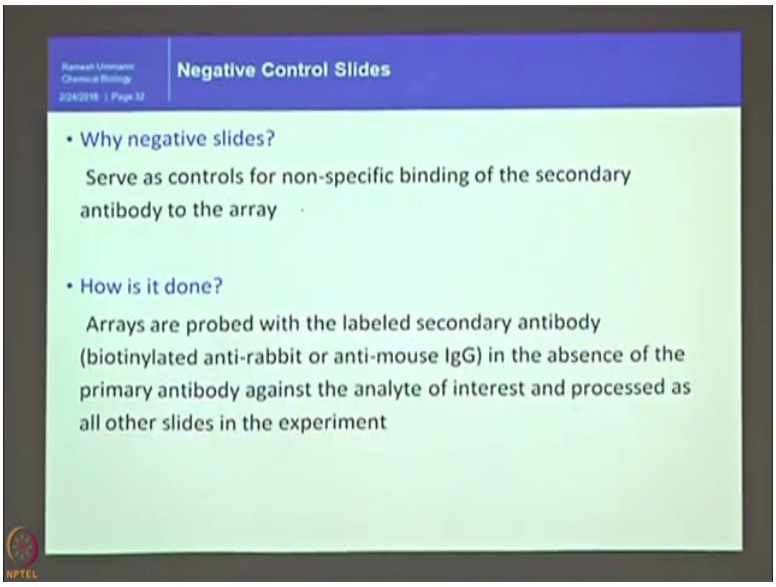
(Refer Slide Time: 04:56)



So, we stain the slant with fast green FCF; this is called FCF. So, this will directly bind to the protein and we can screen sorry we can scan the slide at 800 channel in infrared imaging system. This green color indicates the total protein. We can quantify this same as like antibody signal quantified and we can see that.

So, we can already see that in this plot, there are 5 or 6 different samples going at different way depending on the time and these signals can be normalized. After normalization once you plot this normalized data raw data, you look like this. To get a more smoother data, once we normalize this with FCF can get a signals like this. You can clearly see that already how nicely the data can be, data can be collected to make it biological sensible.

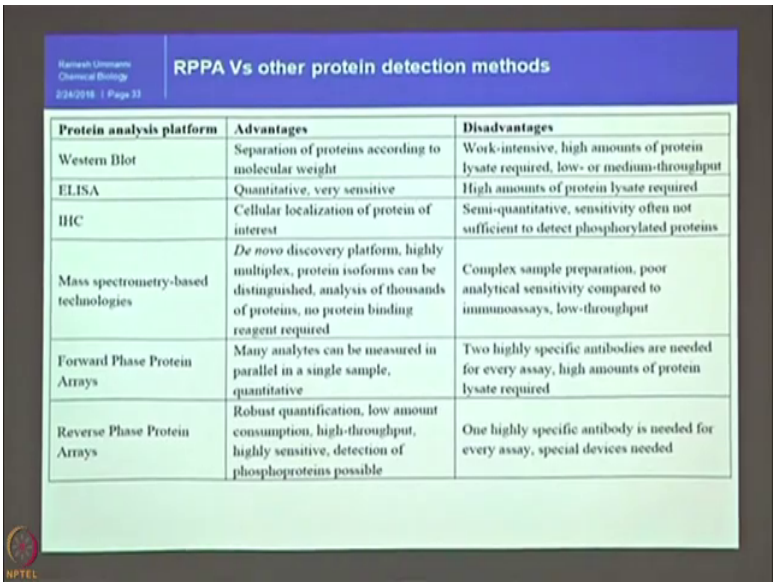
(Refer Slide Time: 05:44)



So, with this and also morning he suggested to have what kind of controls to include and how to plan experiment and all. So, it is important to include negative slides. So, these negative slide server as non specific binding of secondary antibody to the arrays. How this array, every one of the slide out of 50 will be probed with the only secondary antibody and those signals will be deducted from the original signal. And if you see any signal for the secondary antibody alone. So, those antibodies definitely need to be dropped.

There are some antibodies which, there are some secondary antibodies labeled for infrared from different companies behave completely different.

(Refer Slide Time: 06:29)



So, now to summarize here how RPPA is advantages or not advantages over the different methods for detection of proteins. First you look at the western blot. Advantages separation of proteins according to molecular weight. So, now, you know the what molecular weight your target of interest is, those Santa Cruz will give ten different list, but you know which molecular weight you should look for.

The sorry ELISA, but in the disadvantage is very work intensive, high amounts high amount of protein is required and it is very low throughput. Maximum you can do medium throughput, but which is not sufficient to analyze thousands of samples. And ELISA which is quantitative, very sensitive, but it requires high amount of lysate. Then IHC is a cellular localization of protein of interest. It is also semi quantitative and sensitivity often not sufficient to detect phosphorylated proteins.

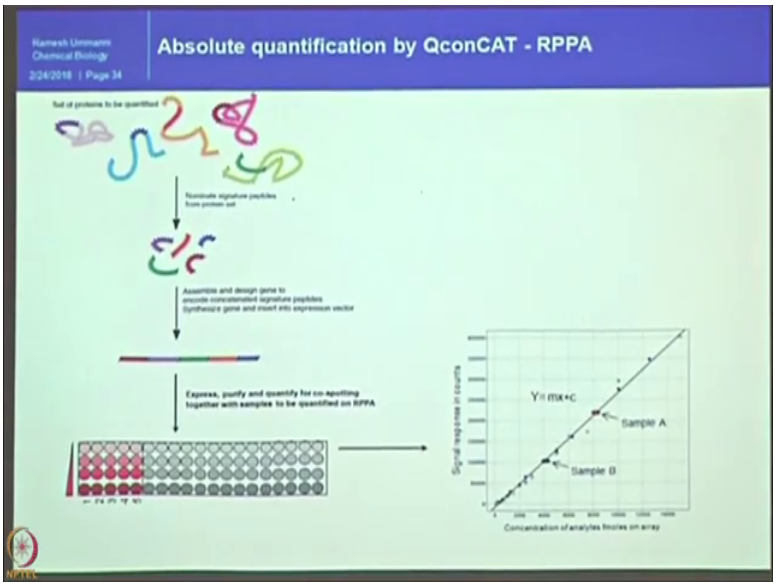
Only very very few copies of phosphorylated proteins will be there in the cell and then it is very difficult to look at them in the IHC. And mass spectrometry based technologies. I am not saying that is not good, in fact it is a De novo discovery platform, but highly multiplexing can be done, protein isoforms can be distinguished, analysis of thousands of proteins and no protein binding regent is required. It is a just direct injection to the sample.

This is a main advantage, but disadvantages if you look for complex sample preparation, poor analytical sensitivity compared to immunoassays and definitely is a low throughput compared to the array platforms. I will not say that it is very low throughput, say compared to array platforms it is little low throughput. And forward protein arrays, many analytes can be measured in parallel in single sample and it is quantitative. The disadvantage is that two highly specific antibodies are required there also because it is antibody captures the analyte, right. So, capture antibody also should be highly specific.

But reverse phase protein array is robust quantification, low amount of consumption, high throughput, highly sensitive and detection of phosphoproteins. It is not that I am vouching that it will meet all the other disadvantages of the methods mentioned here. It also has a limitation that it has highly, it requires highly specific antibody and every assay need to be optimized and special devices required for this technique as well.

So, this is how the RPPA works and also you might have seen that always I am talking about qualitative. Now mass spectrometry evolved much higher level that even can perform the quantitative absolute quantification, even the number of copies present in per cell and all. So, to some extent here also RPPA is useful to make absolute quantification of proteins this QconCAT approach.

(Refer Slide Time: 09:32)



That means quantitative concavalent protein quantitative approach, that is why it is called as like this. So, let us say for example, you can you wanted to establish QconCAT approach. You select a protein of interest and you selects epitopes recognized by the antibodies.

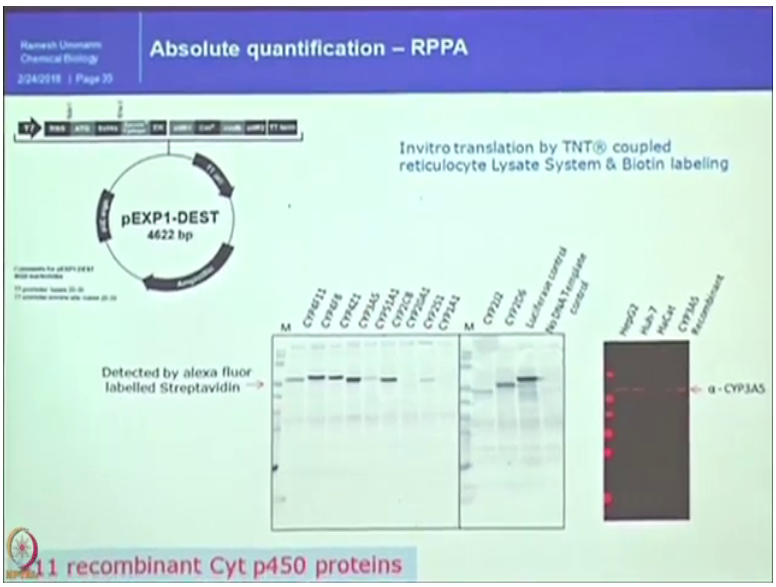
So, here you have a like kind of always when you buy a antibody, but antibody will bind to the certain region of a protein, it will not bind to the entire region of a protein, right. So, you can also develop a method called QconCAT approach by selecting those epitope regions of a proteins and you can express them as a fusion proteins.

So, this definitely useful in all kind of protein array approaches and you can develop a single fragment like this containing different epitopes and those proteins can be printed as a standards here in a concentration dependent manner. And then you can do the RPPA along with unknown samples here then you can make a graph like this, concentration of analytes on array

and the signals on the y axis and you can also extrapolate the signals of samples for example, A and B on this plot.

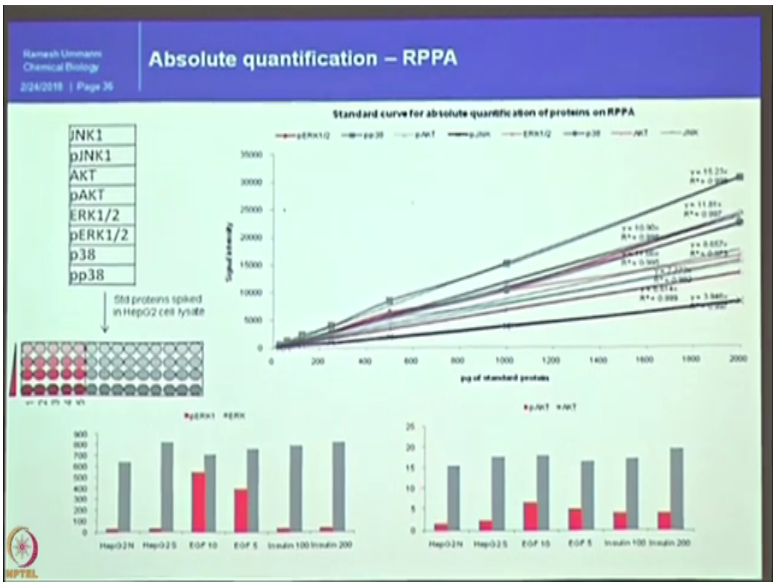
And you can extrapolate and calculate the concentration exactly. If you know the number of cells you have taken to prepare the sample B, you can also say that this many copies present for this many population of cells. This is a way you can quantify.

(Refer Slide Time: 11:02)



I will show one example, in this case at least for NASH project, nonalcoholic steatohepatitis project; we try to express all 11 recombinant cytochrome p 450 proteins as a fusion proteins and using this vector system, we develop the proteins have purified them and we check the antibody specificity and use them as a standard proteins to quantify the cytochrome p 450 protein levels in a certain population of cells. Say, if 1 million cells contain how many copies of cytochrome p 450 proteins like a sip, these eleven different recombinant proteins.

(Refer Slide Time: 11:40)



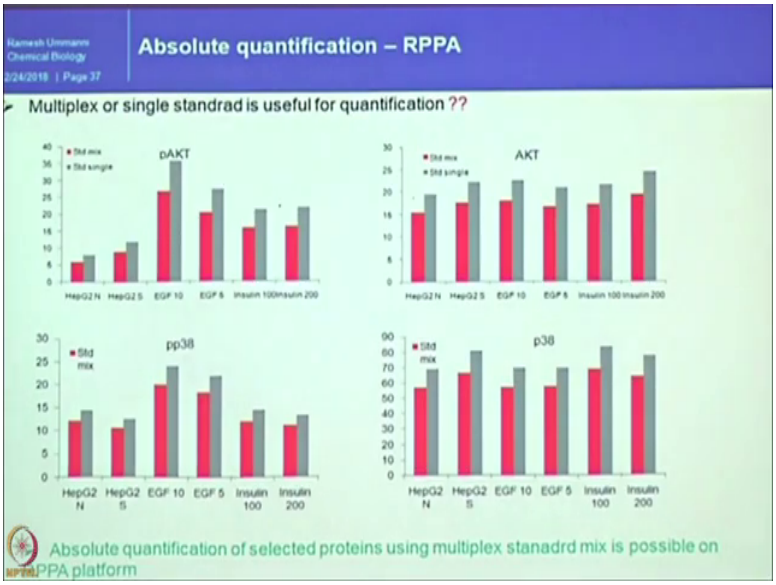
And also, now a days it is very very important to do quantitative approach for phosphoprotein. For example, if you take AKT protein, not all AKT copies are phosphorylated when it gets activated. Only certain population or certain copies of AKT proteins get phosphorylate. How can we do that?

So, what we did is we have expressed up purified express the targets of interest in this case JNK, AKT, ERK and p 38 expressed purified and in vitro translated for protein modifications, sorry in vitro phosphorylated for protein modifications using the specific kinases. And those phosphorylated proteins we spotted in the concentration dependent manner and then later on we did the lysis.

And we can show that phosphorylated proteins on unphosphorylated protein detection, the linearity range and above point 9. And then we looked at the signals detected here, different

cell lines and different stimulated samples, we can nicely detect total values are like this and up regulated down regulated copies are mentioned here. In the similar way for phospho AKT and AKT here. I am sorry the scales are different. Here it is into 10, here is a total number is shown here.

(Refer Slide Time: 13:01)

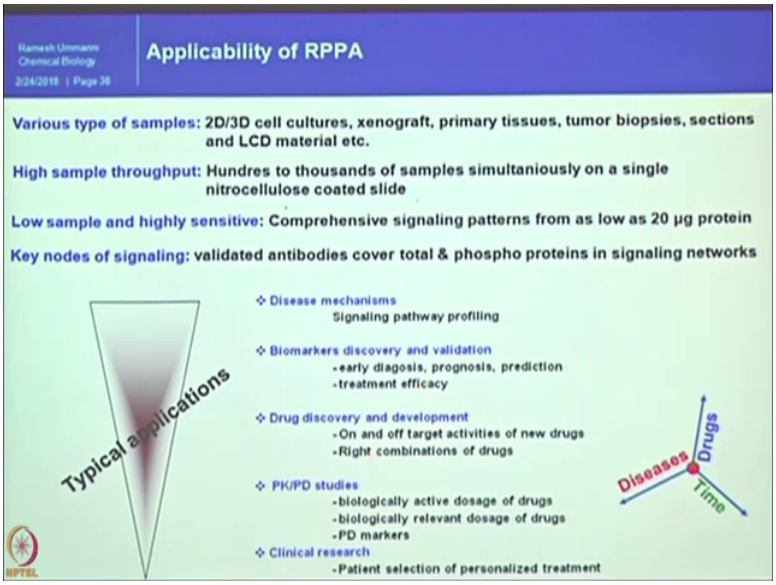


So, in this way one we can perform. The other one is that absolute quantification can be done in multiplexing manner as well. So, on the previous experiment we have spotted separately then we mix all of them, spotted as single spots and a concentration dependent manner. And then when we will look is a single spotting is important or the can be multiplexer is possible.

So, you can also already see that a little bit signal is down, but the behavior is same for all the different proteins. So, you can see the power of RPPA is reached a different level that you can

also use this method for absolute quantification of a particular protein of interest in a larger population of samples, larger set of samples.

(Refer Slide Time: 13:48)



So, now just imagine that this is a technique in your hands, how you imagine to apply into your own area of research is up to you. So, to highlight some of them you can say that 2D or 3D, what kind of samples can be analyzed; 2D or 3D cell cultures, xenograft, primary tissues, tumor biopsies, tissue sections, LCD materials, laser capture micro detection materials is definitely high throughput. This low sample comp comprehensive signaling pattern from as low as 20 microgram protein which is not even sufficient to have one protein in one house keeping gene on the western blot approach.

That key notes of signaling validated antibodies to cover total phosphon total and phospho proteins so that you can reconstruct the signaling networks when you stimulated the cells for

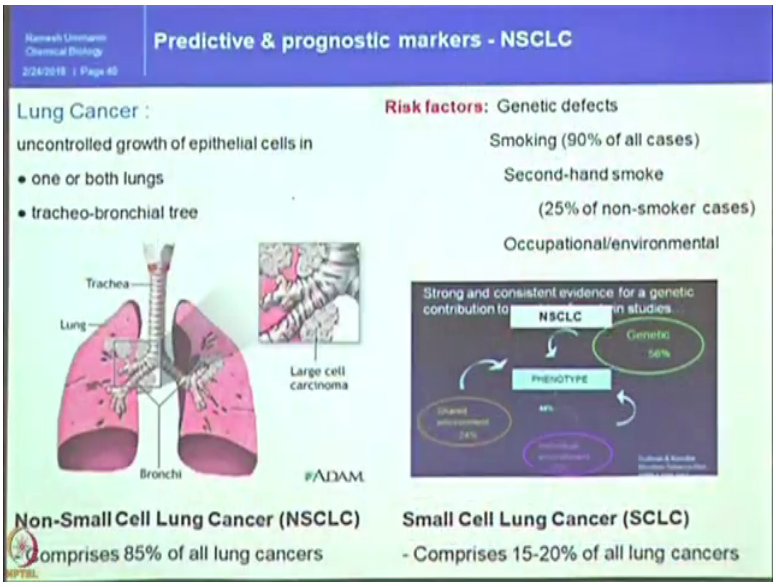
with certain factors or the drugs and so on. So, to typical applications you can use this approach for a disease mechanisms, signaling pathway profiling, biomarker discovery and validation; particularly for array diagnosis, prognosis, prediction treatment and efficacy and drug discovery projects like a on and off target activities of new drugs.

When a new drug is coming into market; definitely now a days regulators are asking what are the off target effects of your drug or what is the target of your drug. Even if we do not you do not know off target effects and right target of this molecule, what are the possible signaling mechanisms or signaling notes targeted by a molecule. This can be definitely achieved by RPPA approach and also right combination of drugs. So, when you looking at the personalized medicine.

So, if you using the cell based models and by reconstructing the signaling pathways, you can choose the right combination of drugs. So, always sometimes combination of combinatory treatment are more effective than the single treatments, ok. So, PKPD studies, you can identify the biological active dosage of drugs because some certain signaling pathway should be shut down or shut on. So, what is the concentration of drug is required to switch on switch off. So, our reading or the end point reading read out is the switching on switch off by using the drug molecule.

This switch on switch off signaling pathways can be definitely done by RPPA in a high throughput fashion. So, clinical research definitely patient selection for phosphorylation treatment I will touch up on later on. So, using this technique we have done different project, at least I have done this different projects.

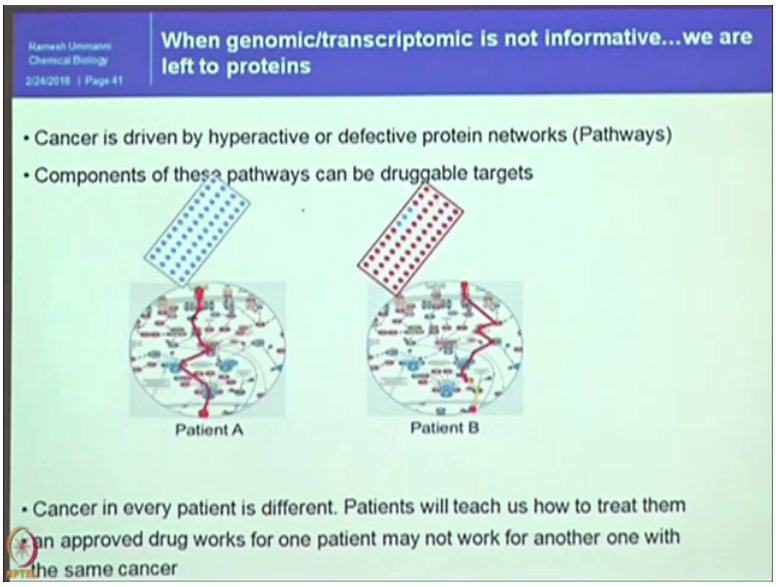
(Refer Slide Time: 16:33)



So, I will touch only non small cell lung cancer and prostate biomarker a little bit. So, the lung cancer I do not want to go more details into the lung cancer and it definitely has two different types; non small cell lung cancer and small cell lung cancer.

Non small cell lung cancer about 85 percent; only 15 to 20 percent are small cell type, but today both cancers are difficult to treat. Though having lot of drugs, it is very difficult to treat and still mortality rate is high and due to the several reasons, that the different sub types of non small cell lung cancer are there in the patient. Now in the clinic, even doctors are not able to segregate them before starting the treatment. So, what kind of drug is good to give to them and all.

(Refer Slide Time: 17:15)

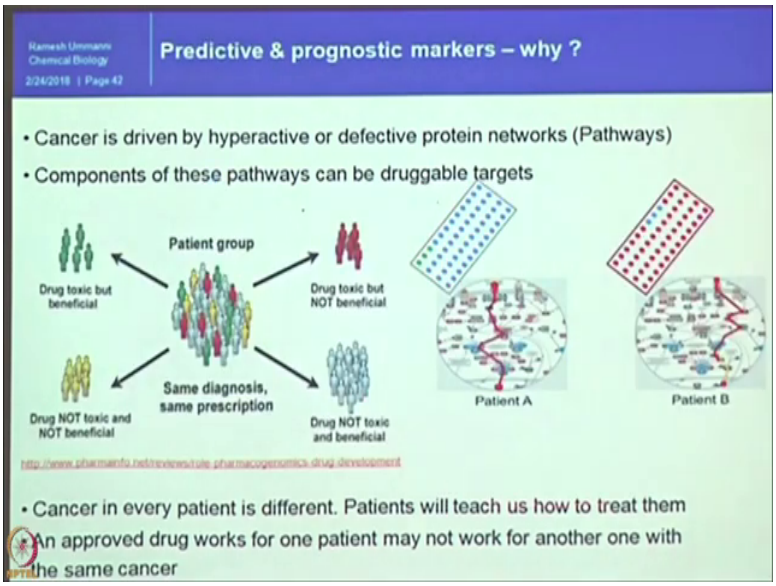


So, keeping that in mind at least we try to do some good experiment on how we can separate the patients based on their molecular subtypes using RPPA approach. I will show you one or two examples why it is required. So, when genomic or transcriptomic is not informative, we are only left to proteins. So, when doctors do not have any genomic information or molecular diagnostic information; by analyzing the proteins, they can take a decision what kind of drugs can be given to the patient.

So, cancer in every patient is different, definitely patients will teach the doctors how what kind of treatment they should receive. In this case patient A and patient B, you can see that red line is passing differently in two different diagrams, but both of them are lung cancers assume in that way. But the signaling pathways are responsible for non small cell lung cancer are different.

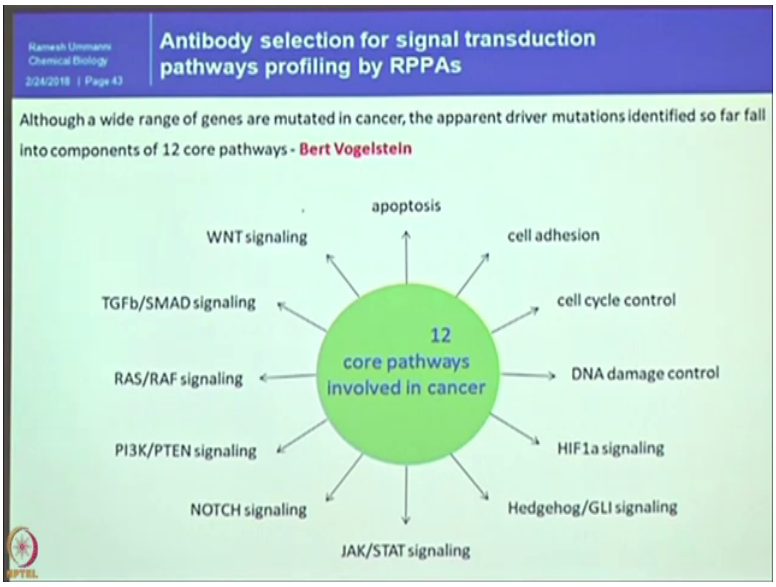
But how doctor knows this? Unless until he analyses the signaling pathways responsible lung cancer only he will be able to understand, how we did it this in this case.

(Refer Slide Time: 18:21)



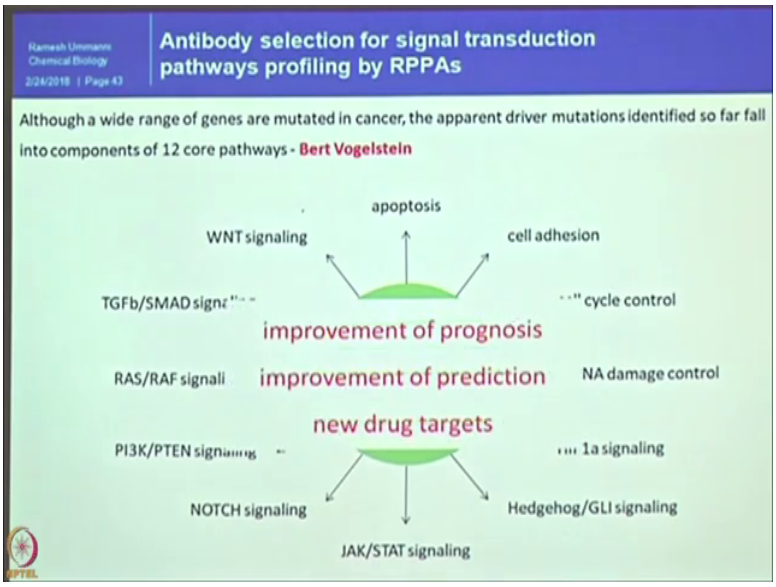
So, again predictive and prognostic markers are required whether the disease is getting cured or not cured.

(Refer Slide Time: 18:28)



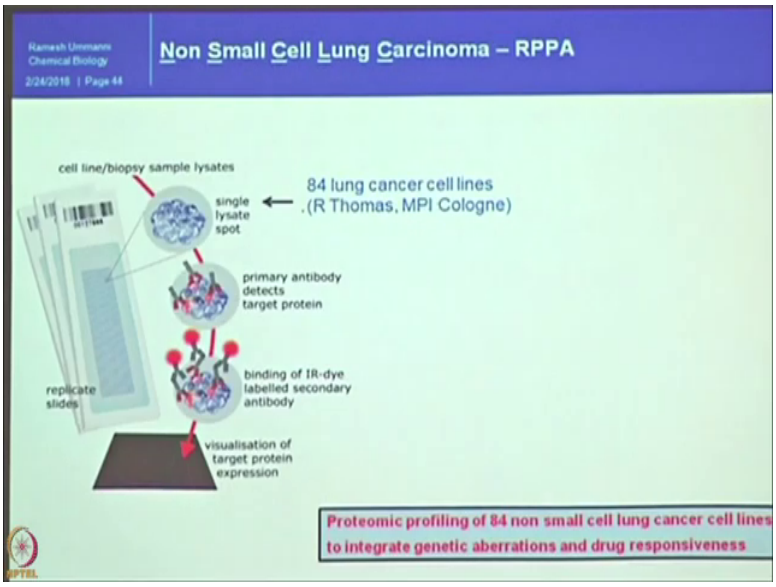
So, I will show with some examples. Again as I said in the first couple of slides antibodies selection is very very critical and important to establish this method. Although wide range of genes are mutated in cancer along with lung cancer, the apparent driver mutations identified so far are components of 12 core pathways; that means, all the different mutations which cause lung cancer map into one of the these pathways; that means, if you analyze this pathways activation inactivation in lung cancer patients, based on the activation status of the pathways you can choose the drug to be given to the patient, this was the motto.

(Refer Slide Time: 19:14)



We decided or we arranged a panel of antibodies which are measuring activation status of this pathways to proceed forward.

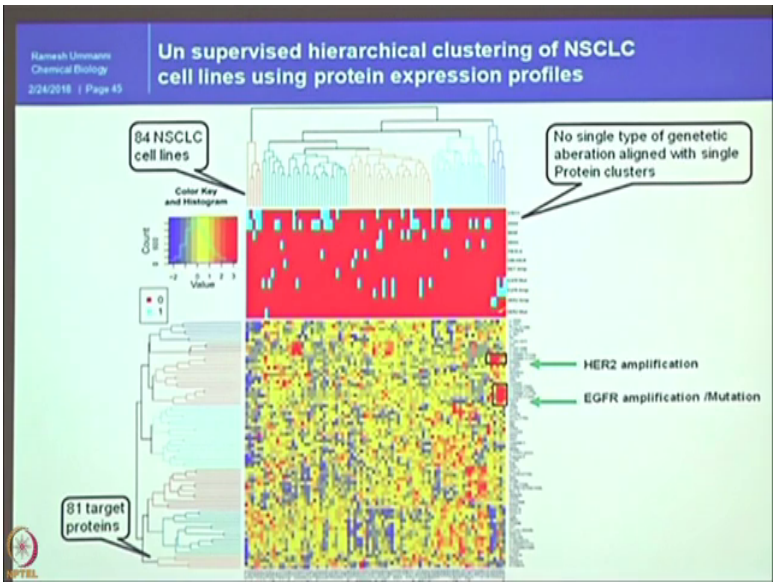
(Refer Slide Time: 19:21)



Of course it is very very difficult to get patient samples with all possible molecular subtypes but one of the group in cologne, Roman Thomas group has established 84 lung cancer cell lines from different molecular subtypes of the patients.

So, we know what kind of mutation they have, all the genomic data, genomic fingerprints are available on those cell lines were established, but what are the different pathways activated inactivated is not known. So, in this case these are the genetic abbreviations which are causing lung cancer today and these 84 cell lines have this 11 different mutations and we perform the protein proteomic profiling again using the 782 antibodies which are from the 12 different pathways in the previous slide, ultimately we get a heat map like this.

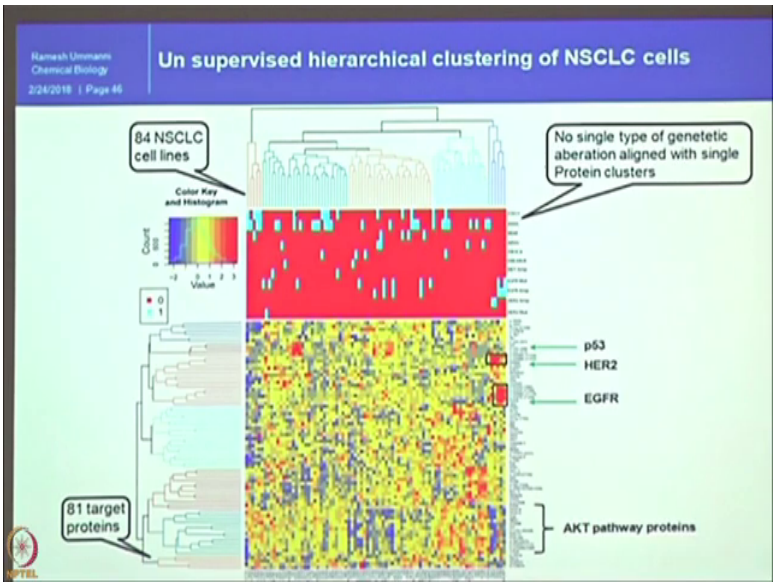
(Refer Slide Time: 20:08)



And you can see that some of the blue areas concentrated, some of the red areas concentrated and on the x axis you can see the cell line and y axis are the different proteins we measured and these proteins are divided into almost 5,6 clusters. All the proteins are divided or expressed as different clusters, here the sion color indicates the presence of the mutation and red color is not presence of the particular mutations, when we try to align this some any of this concentrated areas are not aligning with sion colors.

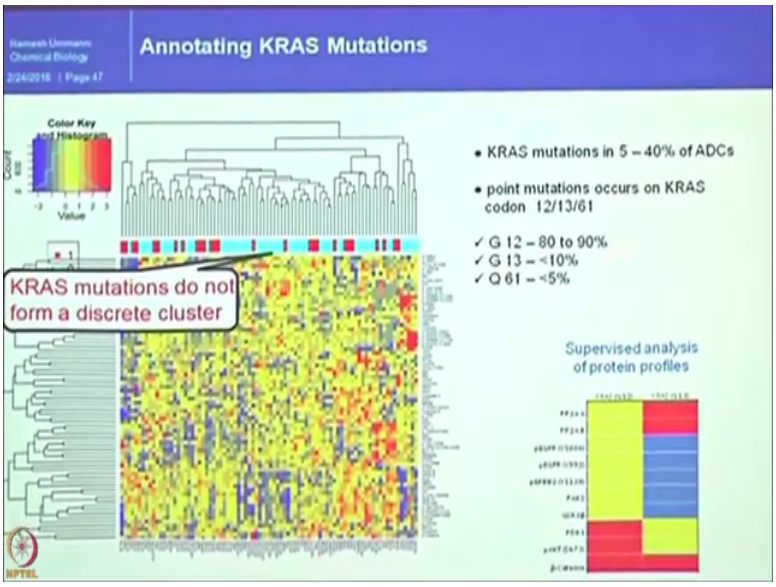
So, wide spread is there you can see that, but here we need to take intelligent decision here. What we have done is did a closer look, here we can see that HER2 amplification and EGFR amplification are mutation signals, this activation of EGFR pathways are nicely aligned with these 3 places here, you do not see any sions here right.

(Refer Slide Time: 21:09)



So, this we looked at again, this is some of the AKT pathway proteins are inversely related to here, the red means up regulated, blue means down regulated, yellow means no change in date. So, this blue color is inversely related inversely proportional to the red color of z here.

(Refer Slide Time: 21:24)

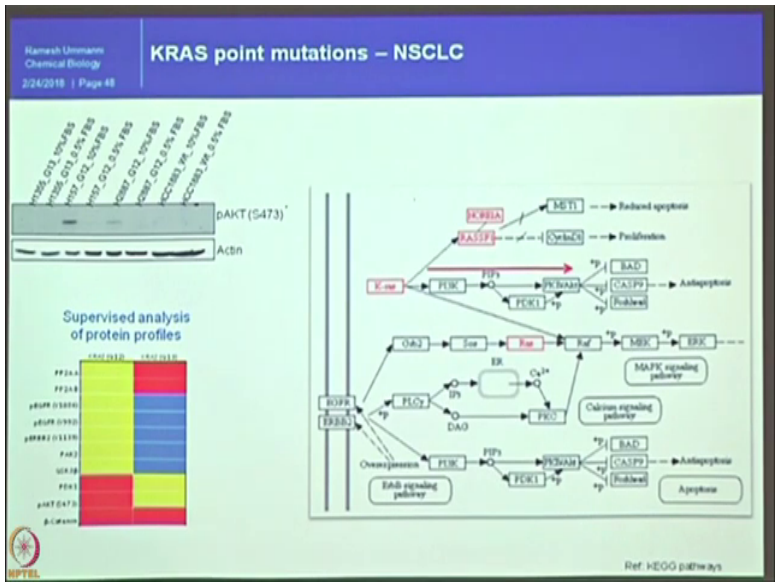


So, then we looked at the in detail, so major cell lines population is having KRAS mutations and also KRAS protein can be mutated in 3 different places. So, usually when a patient is declared that KRAS mutation is present doctor think that x drug can be given, but within the KRAS mutated patients if they are different sub populations, drug works for some set of peoples some set of people will not work.

This can be addressed by RPPA approach how you can see that, I will try to convince you that when you look at a closer look here, these are the different proteins up regulated, blue are down regulated in a G 13 mutation compared to G 12 mutation. There are 3 different mutations, 80 to 90 percent of lung cancer patients with G 12 mutations and only 10 percent or less than 10 percent are with G 13 codon mutation, only less than 5 percent of populations are affected with Q 61 mutation.

So, we try to assign this proteomics signatures to individual mutations we observed in the cell lines here.

(Refer Slide Time: 22:28)

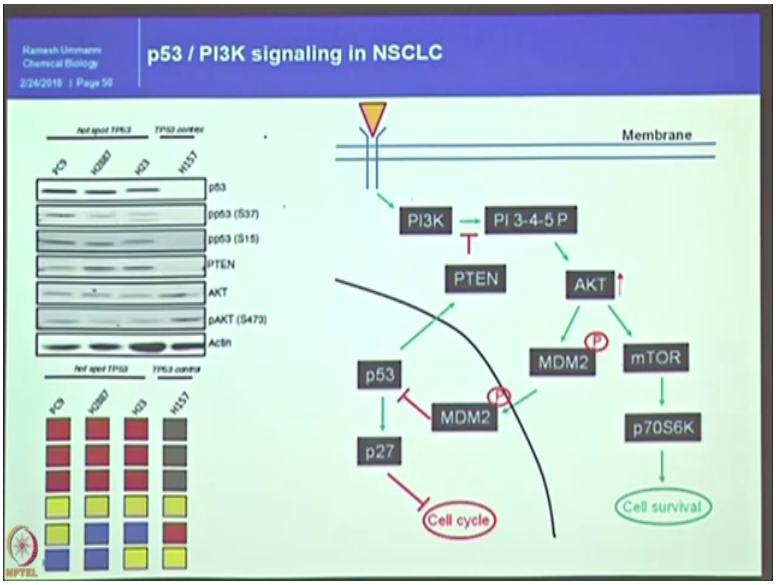


So, with this we did some validation studies. We proposed that, so in the KRAS mutations patients usually treated with RAS Raf inhibitors. So, we also proposed based on this expression pattern that a set of patients with K 13 mutation they should be treated with AKT inhibitors not the RAS inhibitors ok.

So, if the doctor knows that KRAS only, then he will give the RAS Raf RAS pathway inhibitors, definitely set of population with K 13 mutation they will not respond here. So, this kind of information need to be miniaturized in the form of arrays and bring into the clinic to help the patient community here. So, in the similar way we also did for closer look for another

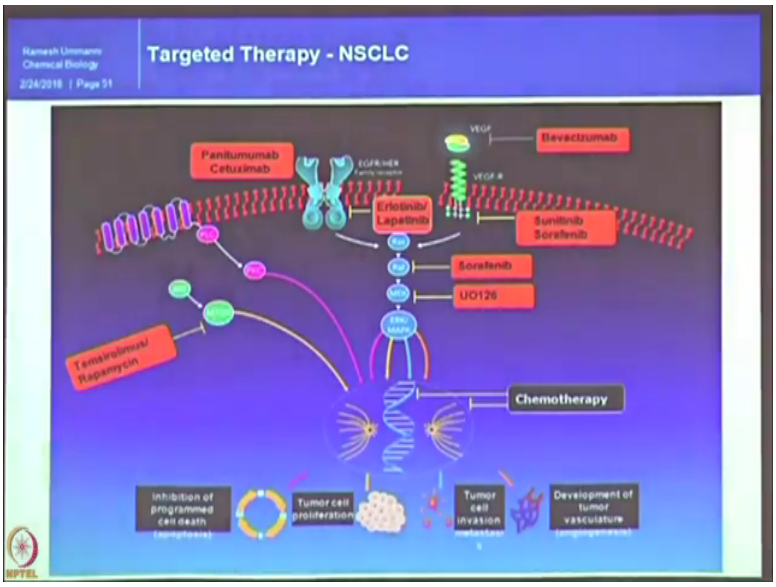
set of small cluster here, you can see that this a kind of small cluster with red patches here and here.

(Refer Slide Time: 23:23)



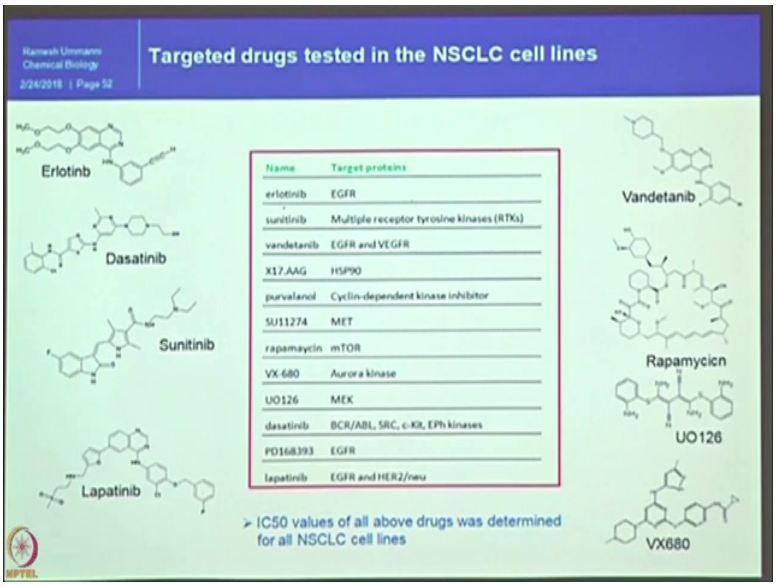
This is basically the p 53 pathway and then we looked at the p 53 hot spots, this kind of pathway we reconstructed so which was not known for p 53 patients; that means, some of the p 53 patients were never receiving the AKT inhibitors earlier, they were always treated in a different manner. So, those p 53 patients with activation status of AKT should be treated with AKT inhibitors rather than the other inhibitors, this information was missing earlier.

(Refer Slide Time: 23:51)



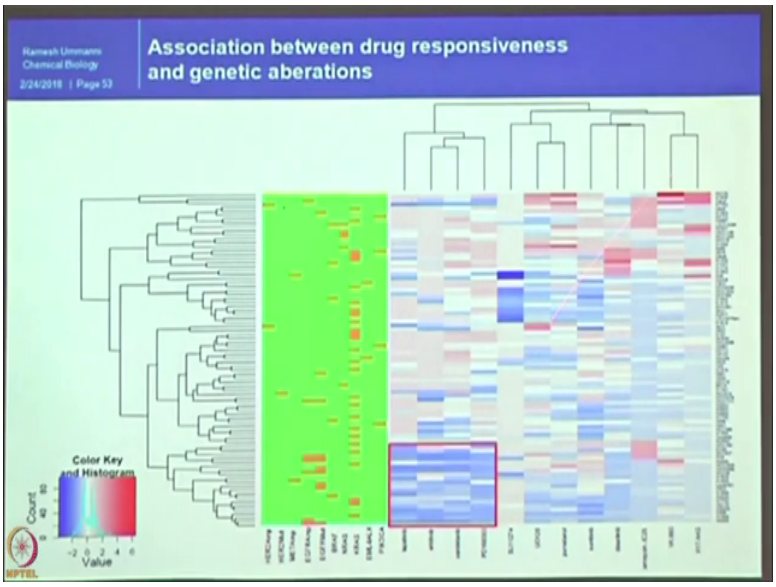
So, this way we can use reverse base protein arrays approach to categorize the patients for treatment approach. And another way is the targeted therapy. So we have a drugs, can we predicts whether the patient is responsive or non-responsive.

(Refer Slide Time: 24:07)



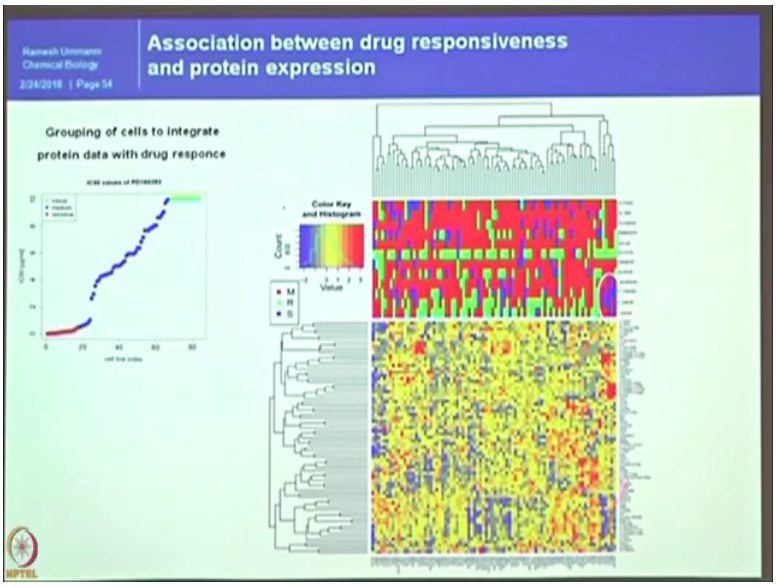
In that case, so these are the different drugs available in the market, we have established the IC 50 values for all the drugs against the 84 cell lines.

(Refer Slide Time: 24:20)



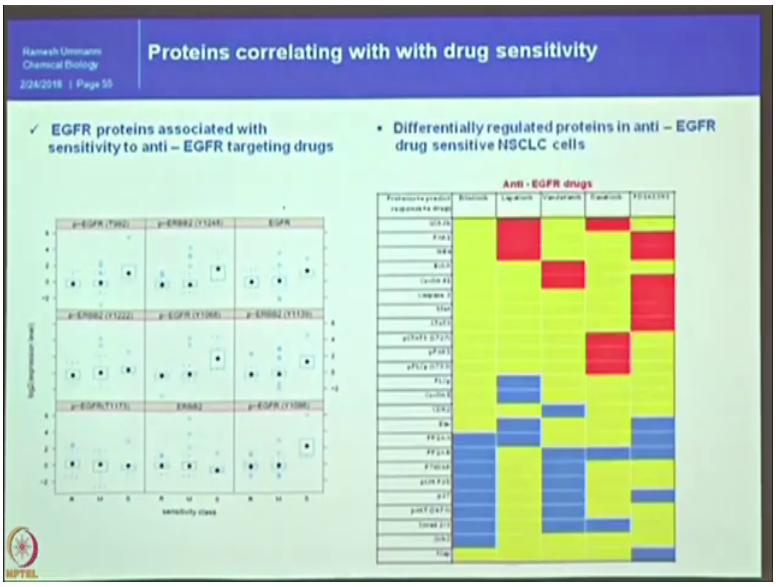
Then we integrated the data using RPPA approach you can see nice cluster here, all these cell lines having the abbreviations for EGFR amplification and EGFR mutations they are responding to the EGFR targeting drugs. That means already established data we can observe in this, whatever the data we generated here is a new; that means, this is a kind of case control for us or can we establish some of the new prediction markers using the other data set.

(Refer Slide Time: 24:51)

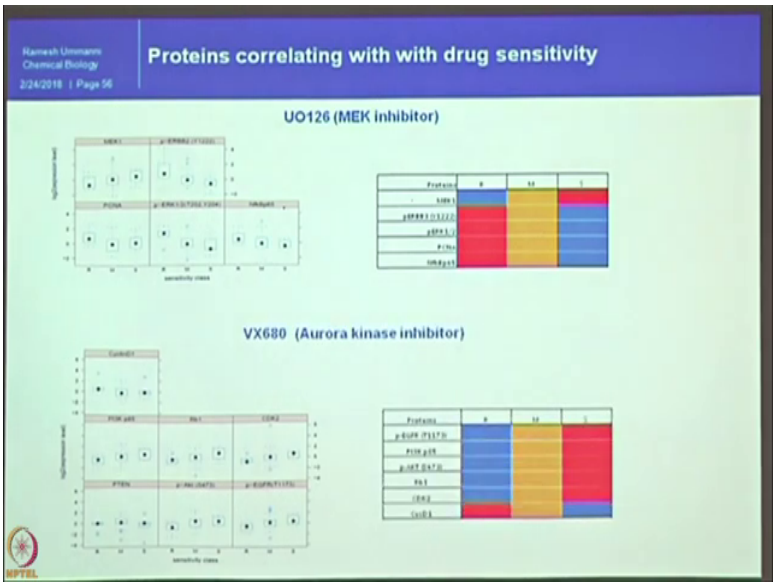


So, using this we also established other data sets in this case, it is nicely aligning here.

(Refer Slide Time: 24:56)

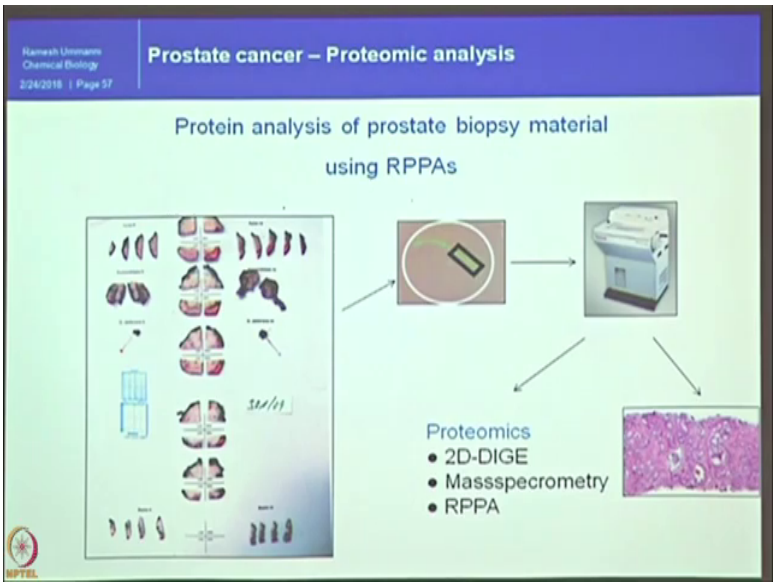


(Refer Slide Time: 24:58)



Further we established a set of markers for all the drugs, we have screened for their IC 50 values against 84 different cancer cell lines.

(Refer Slide Time: 25:07)

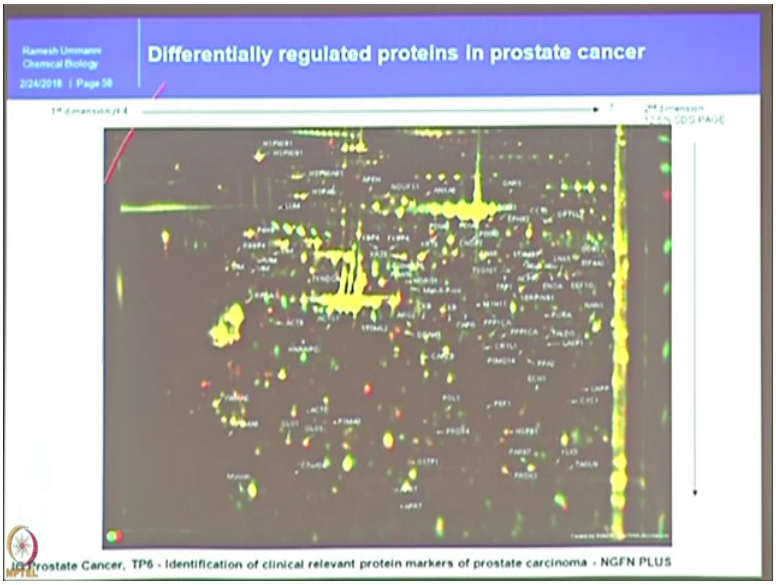


So, this is a way we have established, now last couple of slides I will sum up now. Protein analysis of prostate biopsy material using RPPA approach, then I said in the beginning that any type of sample can be analyzed here. So, the capacity is high through put fashion the applicability of RPPA towards the biomarker discovery approach. One of the study I did during my post got time, so we have got a patient sample and we I sat with the pathologists sections, do the sectioning and get the protein extracted, ultimately end up with a very very small core which can give 50 microgram of sample with a real tumor cell population.

With 50 microgram of sample I should identify the differential express proteins, I should validate them, show the proof that they are really differential express and then write a paper and publish, to get my PhD ok. So, that was a challenging but again here we did at traditional

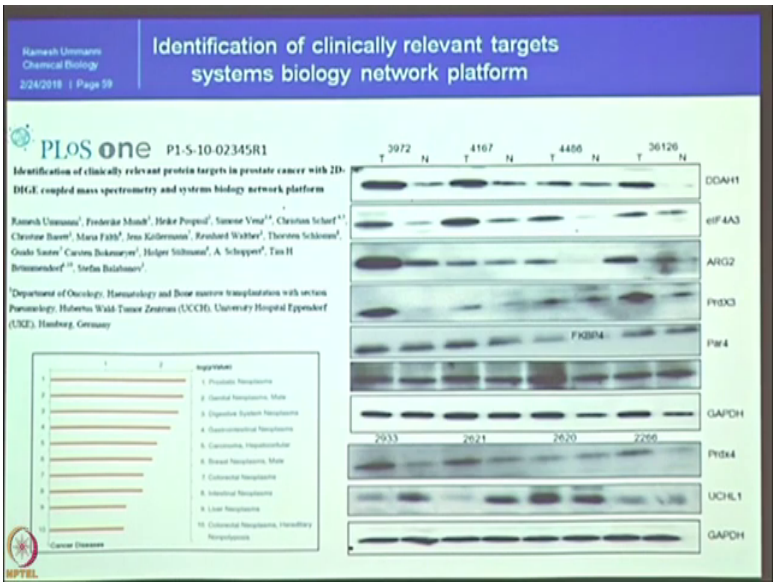
2D-DIGE, now a day's 2D gel is not so encouraging for the people, not very attractive method, those days it was possible to do and did Mass spectrometry, finally we did RPPA.

(Refer Slide Time: 26:17)



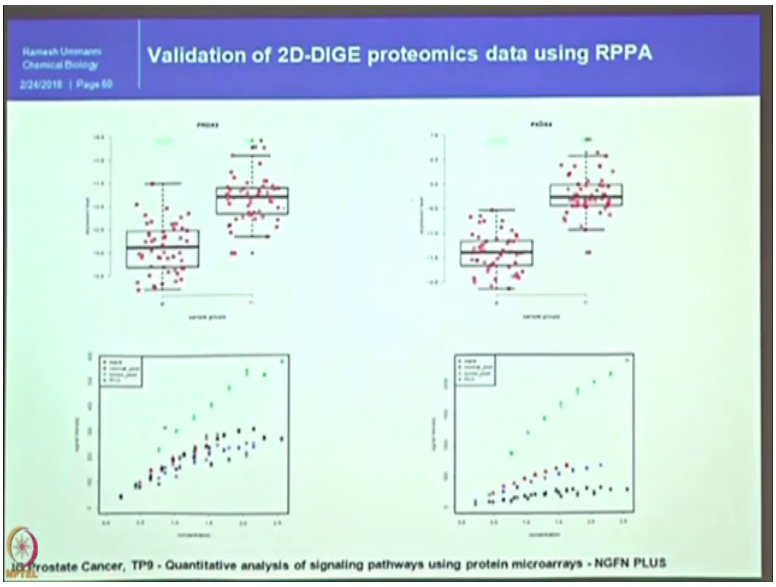
So, these are the different proteins which are differentially expressed, specifically in cancer patient compared to normal or brane prostate hyperplasia in the surrounding tissues. But again this many list is around 112, to validate 112 proteins I do not have a lysate with me, I wanted to analyze and establish the signature from the same set of samples right.

(Refer Slide Time: 26:41)



So, this network platform we built using systems biology based approach and we enrich the marker analysis using the software tools and using the known information these are the different targets at least I have highlighted to proceed forward. This list also not small, so almost 10 proteins are there to go proceed forward, you can see the difference in the tumor and normal very nicely here.

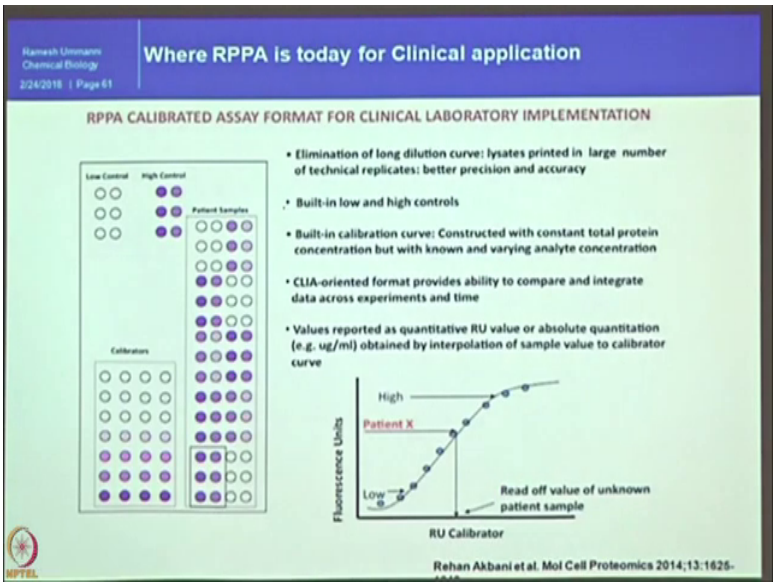
(Refer Slide Time: 27:04)



To proceed forward, then I did this reverse phase protein array approach I could see clearly the difference, for example I am showing pherox redox in 3 and 4 you can see, these each dot represents one patient and it is up regulated in patients samples compared to the normal or control samples.

And also we can see the signal dynamics depending on the serial dilutions and so on. So, based on this approach, we highlighted we could identify that these are the different proteins are really differentially regulated in cancer tissue compared to the normal tissue. From 50 gram protein we ran 2D gel which obviously requires high quantity of protein, we performed a validation of 10 proteins and still 2, 3 micrograms left over in the tube.

(Refer Slide Time: 27:51)



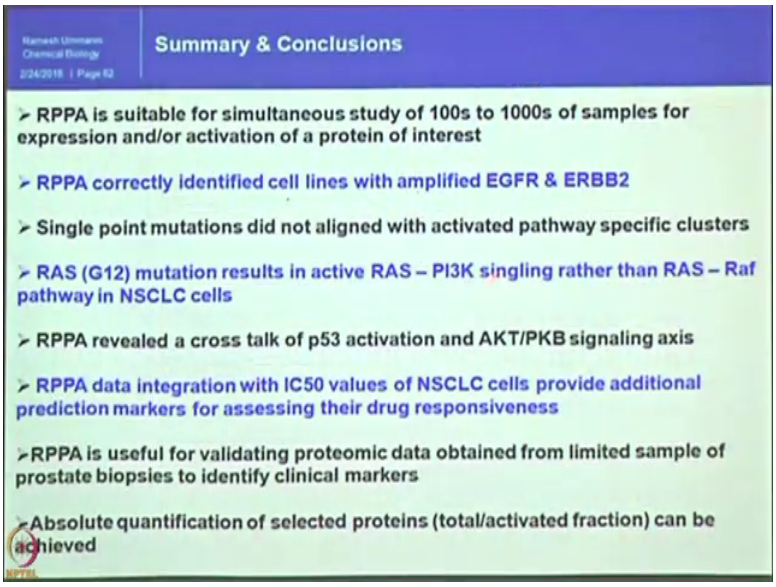
So, this is the power of RPPA we could use. Now where are we standing today and as I said that, that time not many people were accepting that it can go into clinic and all. So, now in 2014, RPPA (Refer Time: 28:02) meeting was there, people were talking about how can we bring it. So, here once we know the level, cut off level, let us say for example blood glucose level we know, it is 80 to 120, below 80 is like a hypoglycemic then above 120 hyperglycemic.

So, in that manner so if you know the levels, we could spot some of the known protein, known concentrations as a low control and high controls and here we can have them as a calibrators based on the concentration dependent, we can construct a plot like this. The moment if we get a patient data in this panel and we can extrapolate on the graph and you can say that what is the quantity, if this quantity is below or between the low and high or above

high or below low, then we can decide that so this patient is having x problem or y problem and so on.

So, now people are talking about this kind of miniaturized RPPA platform to bring into the personalized medicine.

(Refer Slide Time: 29:04)

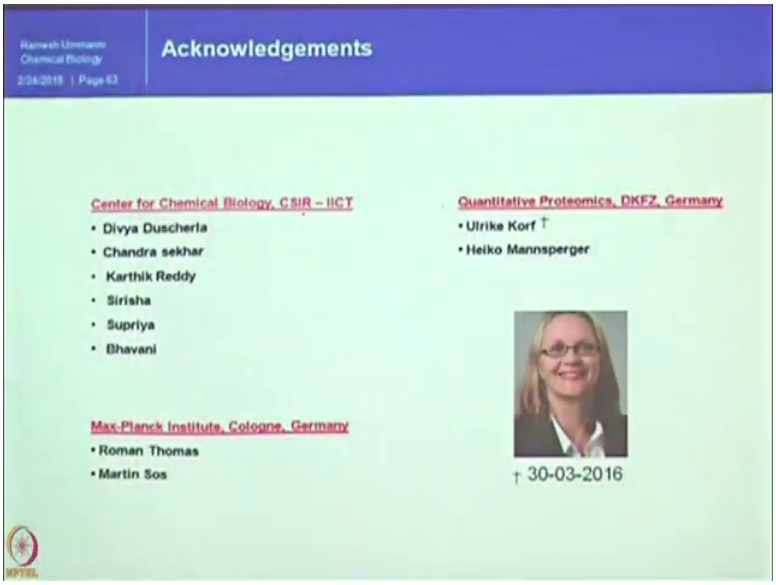


So, with this I will just sum up it is suitable for a simultaneous study of 100s to 1000s of samples for expression and activation of a protein of interest, is correctly identified the cell lines with the EGFR. Single point mutation did not align with activate activated pathways but if the closer look for mutation on G 12 and G 13 is how RAS P K PA 3 K signaling rather than RAS Raf pathways. RPPA revealed a cross talk between p 53 activation AKT which was not known earlier. The RPPA data integration with IC 50 values of provided additional markers.

It is also useful for validation of proteomic data obtained from limited sample and absolute quantification of selected proteins, either it can be total or activated fraction can be achieved per cell in fact. So, this kind of experiments can be done using RPPA approach, but all in all now you might have understood that thus basics steps necessary steps to follow, to printing and then hybridization and all steps are same, only difference between forward arrays and reverse arrays is a here you spot a just screwed lysate. You do not have to follow any separation of samples and so on.

So, on one experiment, on a one slide you can understand or you can estimate quantity of a particular protein of interest across the large number of samples.

(Refer Slide Time: 30:40)



So, this is what. So, with this I will just like to acknowledgement some of the people in my lab and most of the that lab data generated with the help of Ulrike Korf and she unfortunately

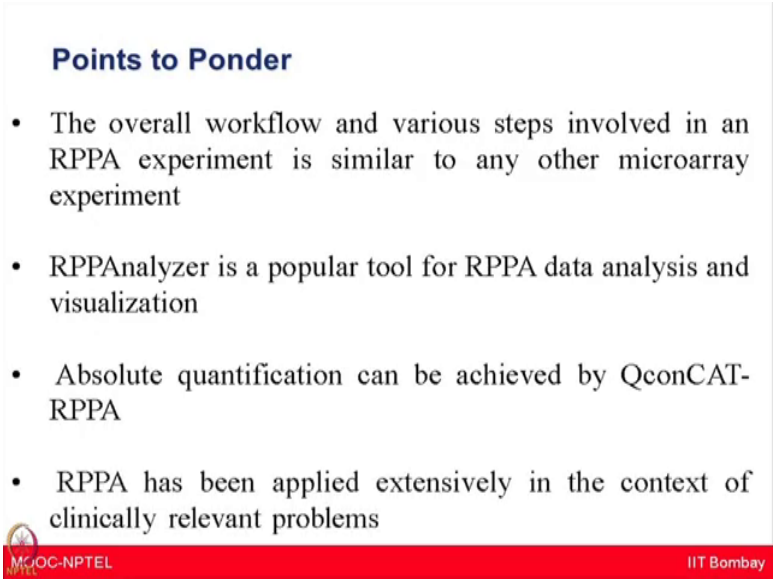
passed away 2 years back. So, she is a one who got me into the reverse phase protein array area in fact. So, with this I thank you all of your patience listening for all most one and half hour probably, I am very open to discuss with you.

So, on in this case again and again I wanted to emphasize that I am may not be able to show you how the slide looks like and all, it is exactly the necessary steps you are following is hundred percent similar, unless until if I tell you that this is reverse forward you will not know it, that is it. So, I hope you will be convinced to do the experiment by virtually imagining the steps I have highlighted you people.

Student: Yes.

Thank you very much.

(Refer Slide Time: 31:40)



So, this now concludes the session on reverse phase protein arrays. I hope you have learnt good workflows and basics of reverse phase protein arrays, setting up the experiment, leading towards the data analysis. By now you know that there are many applications which can be performed on reverse phase protein arrays platform which Dr. Ummanni has explained in context of clinically relevant problems.

I just like to add that now the technologies are really progressing well, and we have you know better printer and the arrays to make these kind of shift platforms. We have sophisticated software tools which can do written tasks better. We have much more high throughput capability.

Nevertheless, the reverse phase protein arrays have been used in clinics from long time. Not as the high throughput technology platforms, but even to test out abundance of a given proteins clinicians have been using it from long time and that also shows the need for having reverse phase protein array based platforms for the clinical applications. So, by knowing that the basics as well as the possible applications, I hope now you are excited to really understand the technology much more detail which can be really helpful if your goal was to deliver something to the clinics for translational research.

Thank you.