Introduction to Proteomics
Dr. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology – Bombay

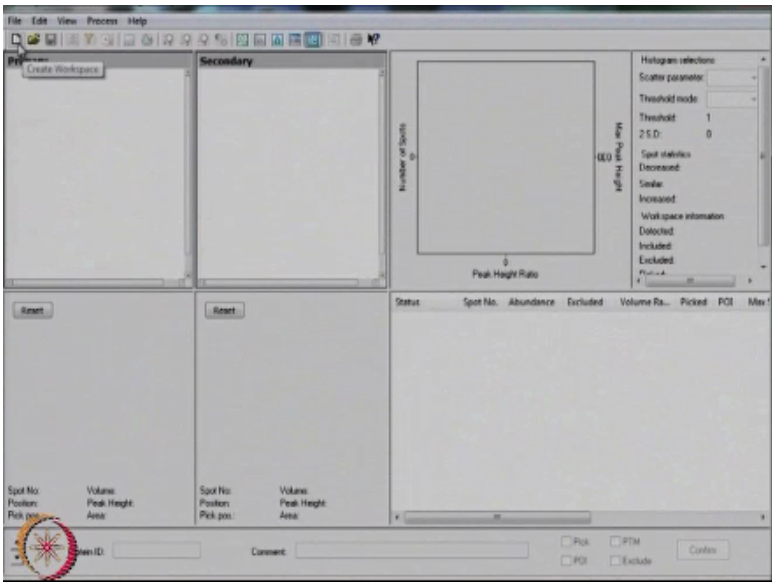
Lecture - 18
2D DIGE: Applications

Dr. Srinivasa from GE healthcare who is going to talk to us about DIGE technology and give us a demonstration on software to perform DIGE Gel analysis.

(Refer Slide Time: 00:38)

The next one is differential in-gel analysis shortly we call as DIA.

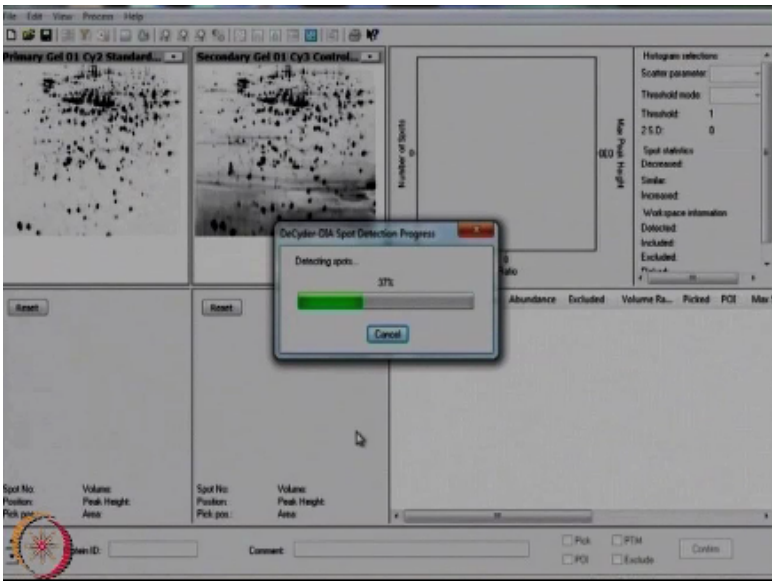
(Refer Slide Time: 00:46)



In this you can see we can create a new project new DIA here like there is an option create

workspace from here you can now it will take you to where we already saved our gels in our database. Now from the database we can select any particular project and from there I am selecting gel number 1 as we saved this one.

(Refer Slide Time: 01:17)



Then now you can see this gel where we have uploaded. So now after uploading here you can process these gels then during process you have to give some number. This is some threshold which you are giving here actually this may be 2000 so that it will take care of background issues also. Then basically in Decyder it co-detection will be (()) (01:50), but I would like to explain you some more about what is co-detection.

This co-detection uses information of all 3 channels and will create the geometrically identical spot boundary for a spot across all the channels that means there are 3 channels Cy2, Cy3, Cy5 out of this 3 channels in Cy2 image it creates a particular volume and the same area can be applicable for the remaining 2 Gels also this is way it works. In this way quantitative and qualitative results are much more accurate than with a single detection.

In DIA each image is co-detected with its internal control producing 2 images pairs the ratio of standard sample is calculated further or the ratio of standard sample is calculated for each protein in each image. So as we see here these are all the number of spots which has been detected in which there are you can see some red color spots here this are all down regulated and compare with control with treated.

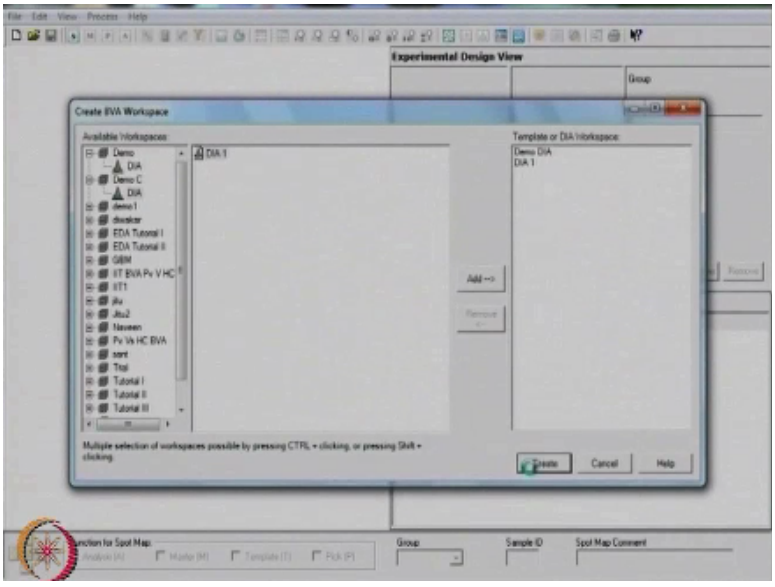
And this blue color spots where you can see these are all up regulated spots when comparing

with control and treated in between this there are some blue color spots these are all similarly regulated. So this is what we can see in DIA. Now you can go through each and individual spot and you can see the 3D view of that particular spot. If suppose we can select anyone it can go to one by one and you can see whether it is exactly spot or some background or if it is a background you have to remove that.

So suppose this is the background, so there is no spot at all still it is detected some background. So you can exclude it from here by clicking then confirm it. So this protein has been removed from the gels so the same way you can go each and individual then exclude it and confirm it. By this way you can check all spots and you can have more accurate data with you like I say you can see the how accuracy would be there like the spot detection.

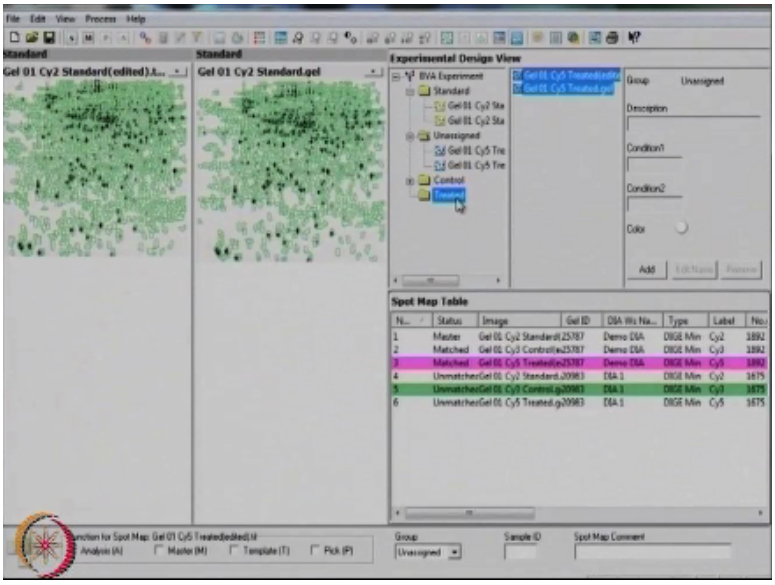
Now you can see this is what we will get in DIA. Thus DIA creation we finished. This is the BVA. BVA is nothing, but biological variation analysis. One of the internal standard image is selected as a master image and all internal standard image matched into this samples, standard, spot ratio for each protein, each sample than compared giving T test value fold changes, ANOVA values for each and individual protein.

(Refer Slide Time: 05:16)



How to create to create BVA workspace? Can open the BVA and create a BVA workspace and go to our DIA workspace where we have our DIAs from there you can create actually add it one DIA I have added and this is another DIA I am adding here minimum 2 DIAs we require for the BVA. So we have 2 DIAs here and click on create. So it creates a new BVA for you.

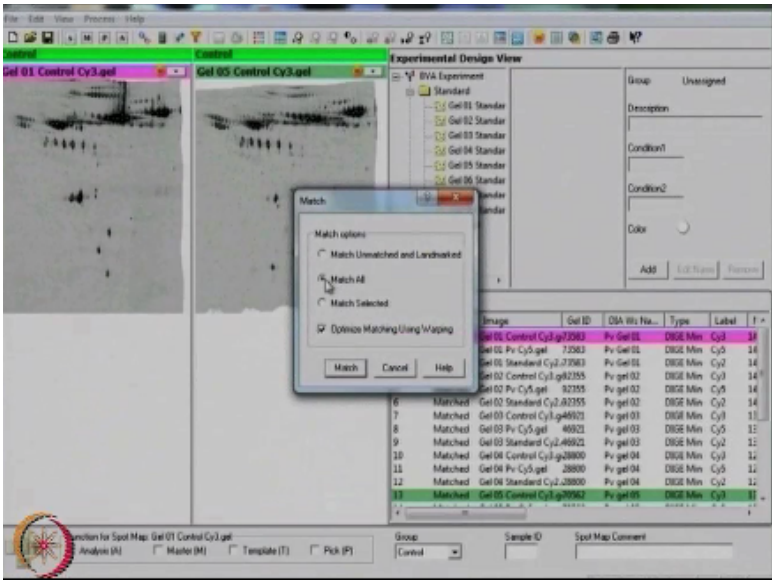
(Refer Slide Time: 05:56)



Okay this is the new BVA from here the first all Cy2 Gels are automatically going to standard folder. There is a standard folder you can see and remaining all gels remains in the unassigned folder where we need to assign these gels as according to the gel type or sample type. Then we just have to click on add option. Now create a group may be control or treated. This first one is the control and apply some color draft, confirm it.

Then another you can create like treated then give some color draft, confirm it. Now we have 2 folders control and treated. So as we have in assign folder both control and treated. This control gels we can transfer into control folder by dragging those images and treated gels we can transfer to treated folder by dragging them.

(Refer Slide Time: 07:28)

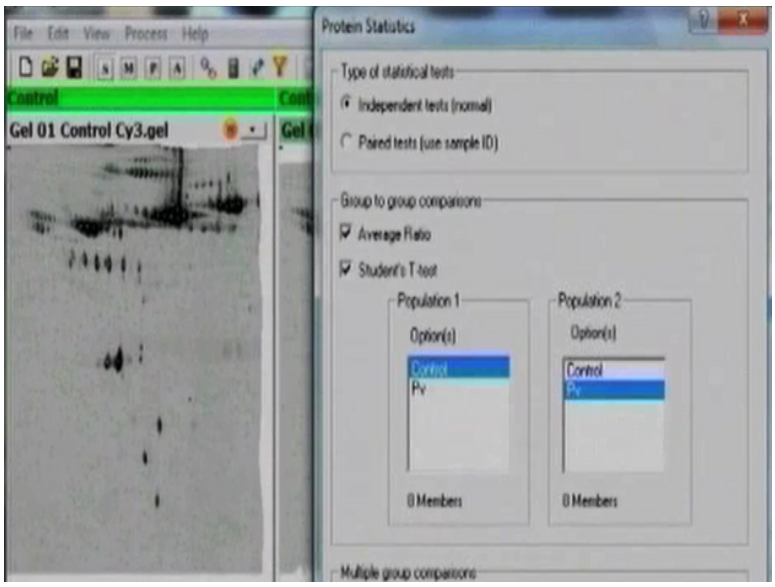


Now we have our images here after shifting the control to control and treated to treated we

have to match all gels just click on match and match all just match. The matching process has been finished. Now as we discussed in out of all standard gels 1 gel selected as a master image as you can see the number 24 gel has been selected as a master and it will compares remaining other gels with this master gel.

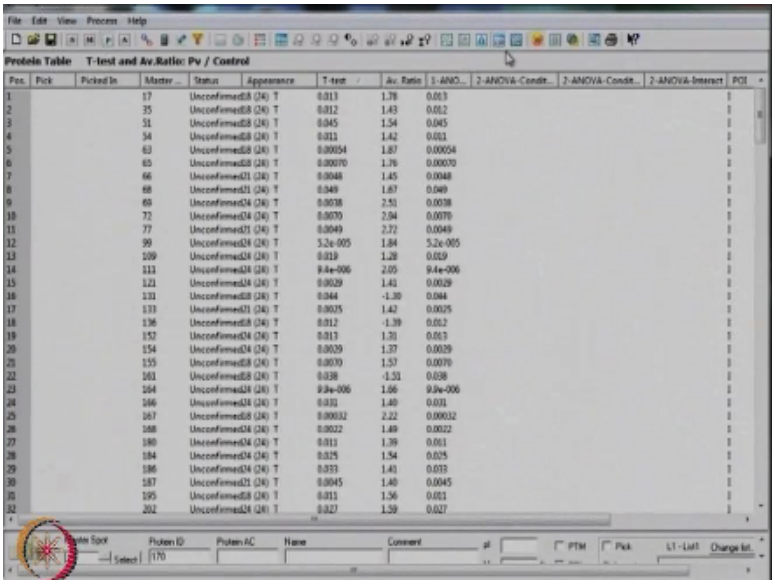
So now we have this comparison data after that. We need to calculate statistical parameters so click on statistical parameter button.

(Refer Slide Time: 08:20)



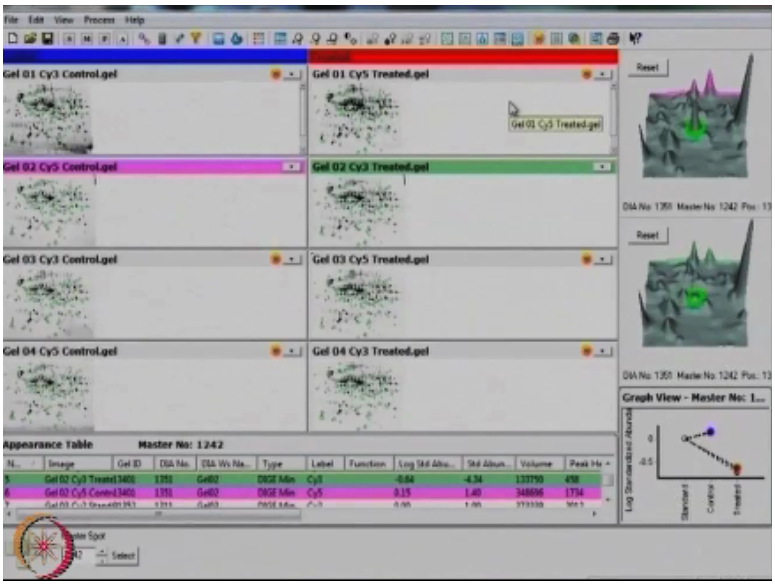
See now we have some paired statistical parameters like independent T test, average ratio, student's T test, one-way ANOVA in between different groups. We are doing in between control and treated. So calculate them so the statistical parameters have been calculated now we can see exact visuals of statistical parameters.

(Refer Slide Time: 08:52)



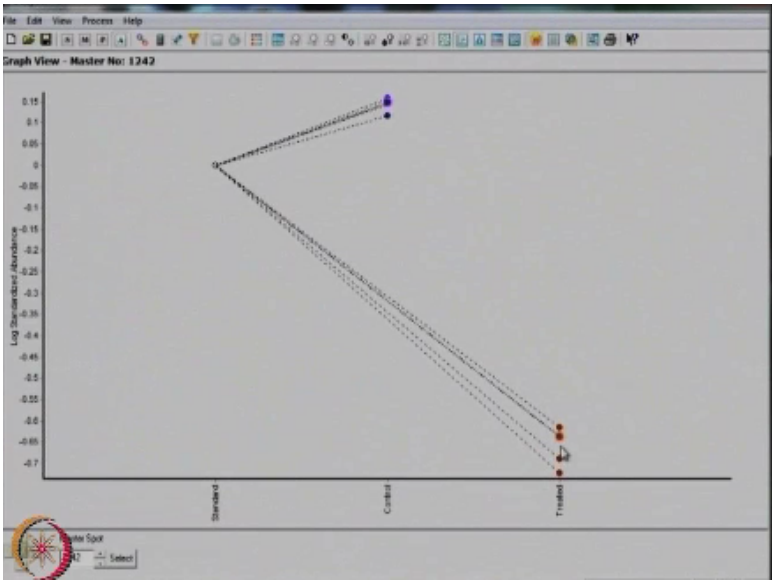
If I can go to the table view here you can see T test, one-way ANOVA this you can see. So we can select from here which are all the statistical significant and which are all not significant.

(Refer Slide Time: 09:10)



After analyzing the statistical data now we can see the complete results here. Here we can see the 4 views like this is the image view.

(Refer Slide Time: 09:22)

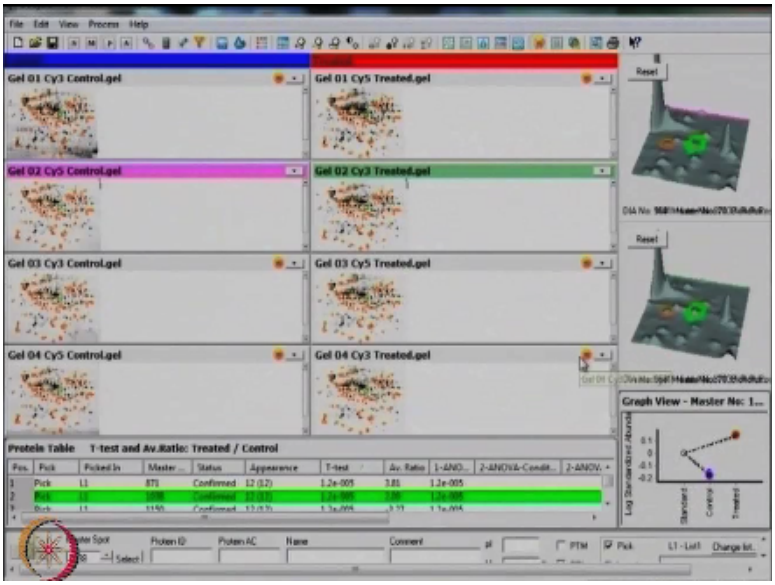


And this is the histogram view. In histogram view we can see clearly particular protein how it is behaving throughout control and treated. If you can see this is the standard gel that means this is the mixture of these control and treated so it is somewhere 0 we can consider this one, then the control is showing up regulation and after giving the particular treatment it is showing the down regulation.

So this kind of data we can see here than in the table view. As we can see the complete protein data where the T test value, average ratio value and ANOVA value this all we can see here in the table view. So the 4 views at a time to see this is the 4 view we can see here. So after this we can filter them according to our interest so select the few parameters which are available restrict to tangles which are spot maps which are present a particular protein should be there and student's T test as well as average ratio.

Then one-way ANOVA value then filter it. So there are 2299 spots are available in this all gels, but 107 proteins only passed all these parameters. So these parameters we can select as a protein of interest and assign pick list so that this protein can be saved in a file this file can given further to our spot picker.

(Refer Slide Time: 11:18)



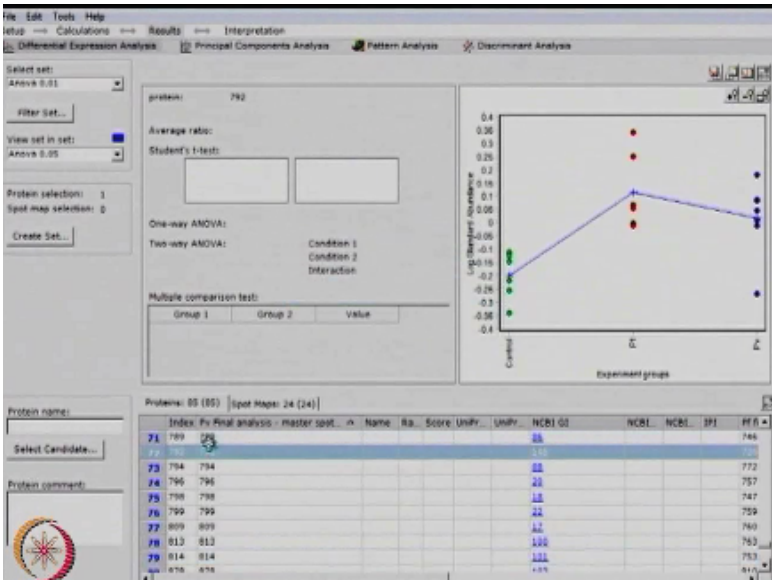
These are all the things we can able to identify very easily in BVA. This is very user friendly. There is no much more manual interference this is what this helps you to analyze your DIGE gels. So can you elaborate on what is EDA or extended data analysis what it can do which we are unable to do it in BVA. So there is the layers here right one is the DIA followed by BVA and then ultimately EDA analysis.

"Professor - student conversation starts" Basically what we can do here is we can compare 2 BVA together there exactly. Here we are talking about a particular disease or a particular set of data only there we can analyze different BVA together in EDA there you will get a majorly differential expression again you will get as well as PCA and discriminate analysis even this kind of statistical data you will get in EDA.

Very shortly I will just show you a briefly regarding that. **"Professor - student conversation ends"** If I understood correctly probably the statistical parameter will be more stringent towards the end in the EDA and you can have some better biological significance and formation from the data set because in lot of clinical data or different type of treatments people like to do several gels and lot of treatments. So your number of sample to be analyzed is very large.

And really obtaining the meaningful information is one of the major challenges in all the proteomics. So I would like to see in a EDA.

(Refer Slide Time: 12:58)

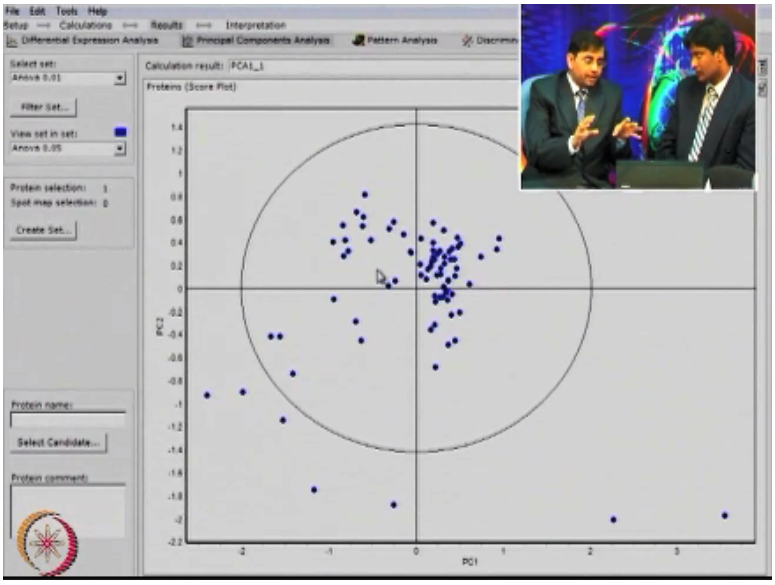


The thing which we can see here is differential expression analysis in which you can see differences in between control as well as 2 experimental data. This is the different treatment were given here you can see how the particular protein is expressing throughout these control as well as (()) (13:18) this kind of things.

And you can see this kind of data for each and individual protein so that from here you can see which one is your interest and which is not of interest. You are actually analyzing the data spot wise now. Here exactly here the spot by spot which we are seeing the number of events the index number which shows there is a master gel from that master gel you can see exactly this number.

This is what which we are seeing here for each individual spot here we can see the results than as well.

(Refer Slide Time: 13:55)

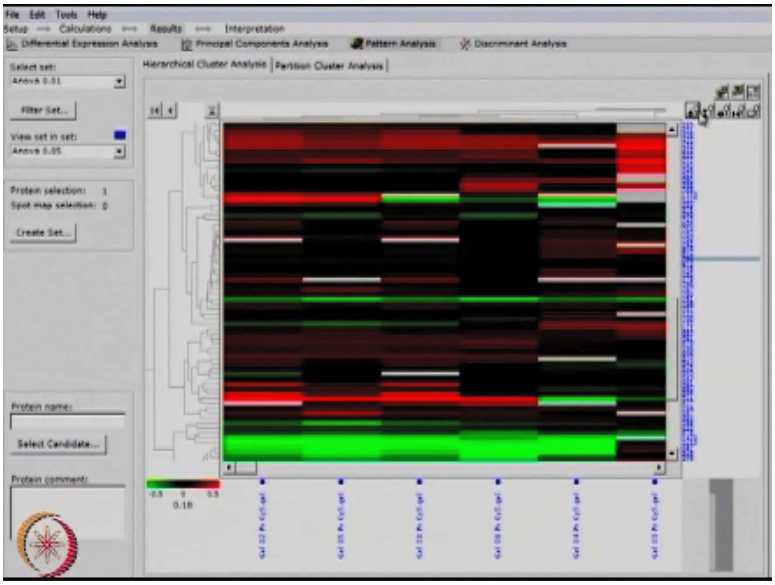


We can see principle component analysis for this data. Here there are almost **18** and proteins. So out of 18 and proteins you can see there are inner the circle there are protein some proteins are present and out-layers are there. The inner circle they are similarly specially if I can say 95% statistically significant there and out-layers which you can see are exactly this can be some non reproducible spots or else what the major things is these are all very highly up regulated or highly down regulated.

So this can be worked as (()) (14:40) also. Then we have to go back to our BVA data and we can check the protein how exactly it is regulating then we can identify the protein and we can use it for further analysis. So this is a powerful statistical parameter by using which you can identify some outlets which could be the potential discriminator between the control and treatment.

And once you identify those proteins then actually you can go back to your original data from the BVA and get all the analysis done so this is very interesting.

(Refer Slide Time: 15:15)



“Professor - student conversation starts”. And next pattern analysis we can see the whole proteome then how these are difference from each other. So this is the hit map of the total 82 proteins which we are taking into consideration then how in which area they are up regulated if you can see the green area exactly we can see is completely down regulated area and the red color portion where you can see that is up regulated portion and the remaining black color which you can see those proteins are similarly regulated this is what which you can see here.

This kind of data will help you to represent your complete whole analysis. **“Professor - student conversation ends.** So Dr. Srinivas it was very useful to have you here and to get an overview of DIGE technology, how people can use Decyder software and analyze the data by using DIA, BVA and EDA and although there was not enough time, but you have given a very good demonstration in a very short time to give a glimpse of the processes involved for doing this analysis.

As well as how different type of statistical parameters can be applied to get some very powerful statistical information from our biological data. So thank you very much for coming here and giving this very good introduction about DIGE technology. So I hope our discussion with Dr. Srinivas was useful and now you can perform this analysis by using a specialized software and obtain some very useful biological information from your data set.

Probably you must appreciate there are lot of meticulous steps are involved in performing these experiments, but at the end this provides a very useful quantitative multiplexing approach to separate proteins and to analyze different type of variations. I hope at the end of

this module and lecture you will be able to perform the gel based proteomics experiment, but please keep in mind these protocols and methods are only giving you a feel for performing these experiments.

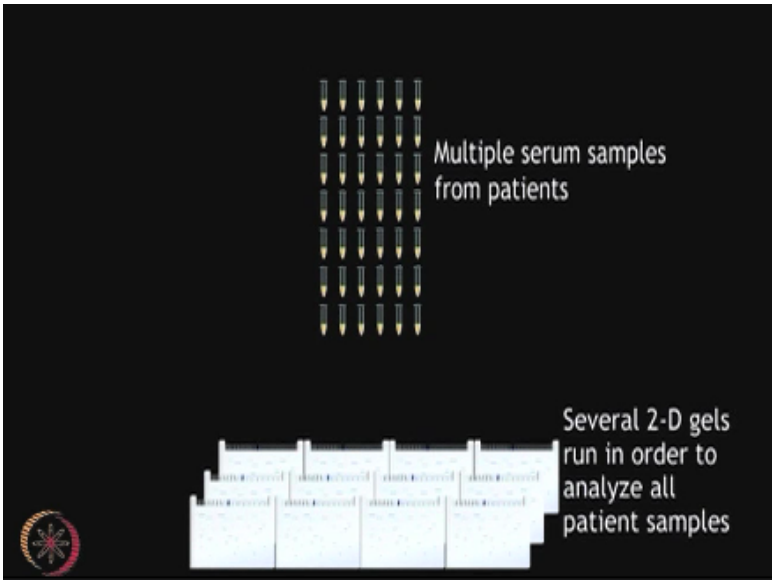
Each experiment, each sample type, each biological questions brings its own unique challenges and depending upon those conditions and your sample type you need to optimize these methods. There is no one technology which can answer all of your question, but it is good idea for you to know that what are different methods which are available for you to use. So I hope by taking this module on gel based proteomics now you are familiar with different type of gel based techniques.

These are only few there are many other methods as available, but these are the most commonly used methods which people are applying in the field of proteomics.

(Refer Slide Time: 18:55)

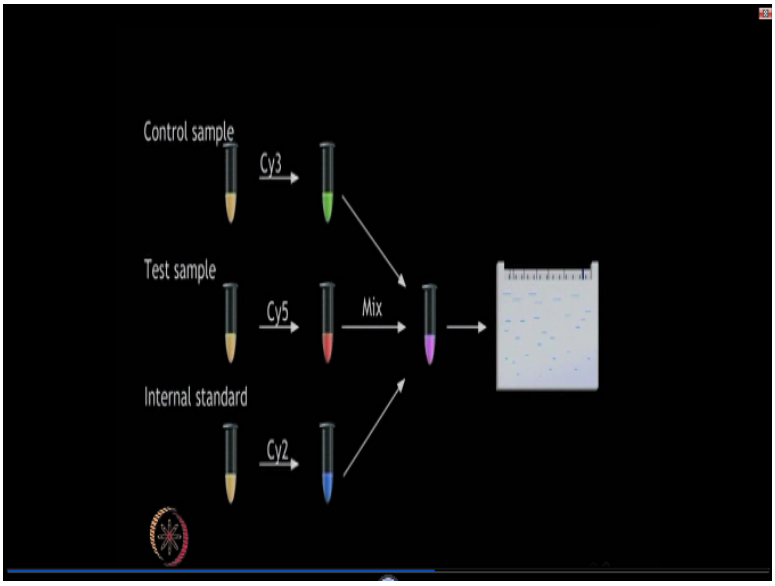
So among 2-DE and DIGE which of these 2 techniques will be better to separate serum protein samples obtained from large number of patients in a clinical trial.

(Refer Slide Time: 19:10)



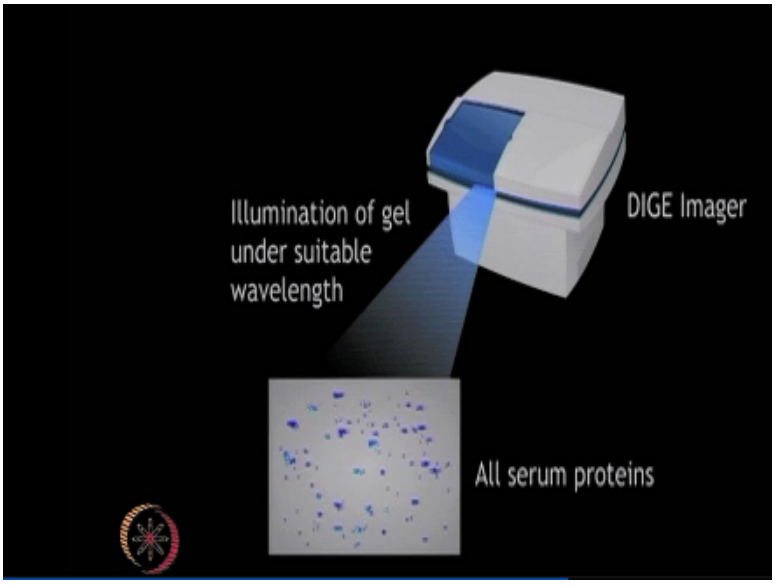
So if you have a multiple serum samples from patients two dimensional electrophoresis although a very useful technique, but it may not be the best option in this case to analyze serum proteins from large number of patients. In this case DIGE will be extremely available tool for analysis of large number of samples simultaneously without having to overcome the problem of gel-to-gel variations.

(Refer Slide Time: 19:49)



In dye gels the controls and test samples can be differentially labeled by using the cyanine DIGE and then run on the single gel. The pool internal standard for DIGE is prepared by mixing equal amounts of all the samples that need to be run on the gel and this prevent the problem of gel-to-gel variations.

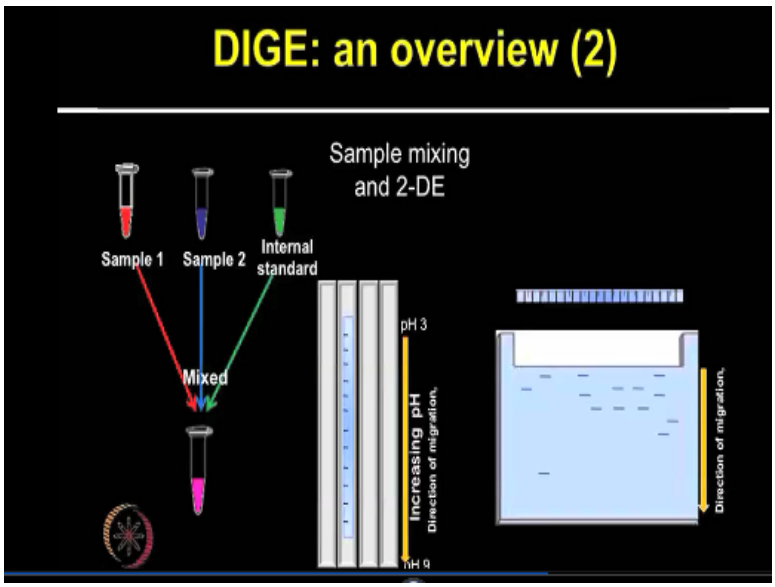
(Refer Slide Time: 20:32)



From the same gel 3 different images can be obtained for Cy2, Cy3 and Cy5. Therefore, there will be no reproducibility issue and various artifacts can be eliminated for the clinical or large number of samples analysis. So the main aim for the development of difference in gel electrophoresis was to overcome the inherently poor reproducibility of conventional two dimensional electrophoresis.

So DIGE is quite sensitive technique with less than one of (()) (21:08) of proteins which can be detected and it can enable the linear detection of very broad dynamic range of the proteins.

(Refer Slide Time: 21:22)



As you can see in this slide the proteins samples let say you have a control and treatment those are labeled with 2 different dyes Cy3 and Cy5, but a small aliquot of both of the samples is mixed together to make a internal pool that internal pool is labeled with (())

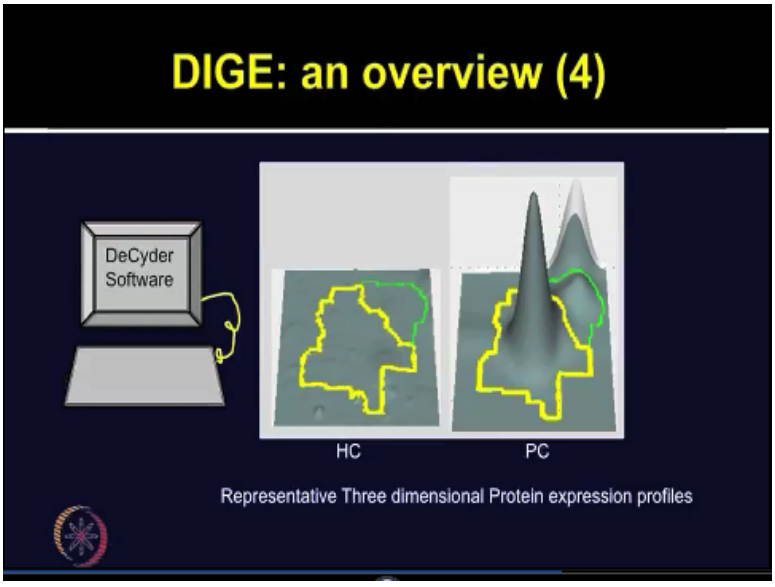
(21:45) Cy2. Now all these 3 protein samples are mixed together in one tube which contains both control treatment as well as the reference spot from the internal pool.

All these protein mixtures are separated in the first dimension on the same strip and then the same gel can be scanned with the 3 different wavelength to obtain the images for the Cy2, Cy3 and Cy5. So in the conventional 2-dimensional electrophoresis the gel-to-gel variations which comes from the Acrylamide polymerization, electrical pH and thermal fluctuations in different gels that can be overcome in the dye gels.

Because all the protein separation is going to happen on the same gel. So all those artifacts can be minimized by using DIGE approach. So in the slide it is shown there is a 3 samples are mixed and then isoelectric focusing is performed from the pool sample. And then this strip is placed on the SDS PAGE gel for the protein separation in the second dimension. So overall DIGE provides very uniform scanning from gel to gel and shows high sensitivity and linear dynamic range of detection for the expression for filing of complex biological samples.

So if your aim to resolve 1000s of proteins and cover comprehensive proteome coverage then DIGE is a very good platform especially if you want to do the comparative or differential proteomic analysis because your gel-to-gel variations and other variations will be minimized and DIGE will provide the very high sensitivity.

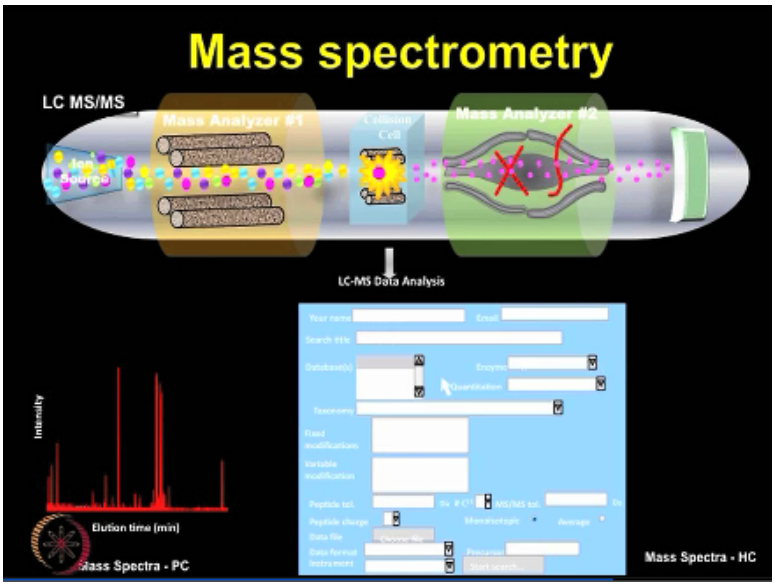
(Refer Slide Time: 24:00)



So once we have run these gels now from the same gel the images can be obtained 3 images of your control and the treatment and these can be analyzed from different software such as

Decyder software and then by looking at 3 dimensional views and the statistical data then these proteins can be considered as interesting for further investigations.

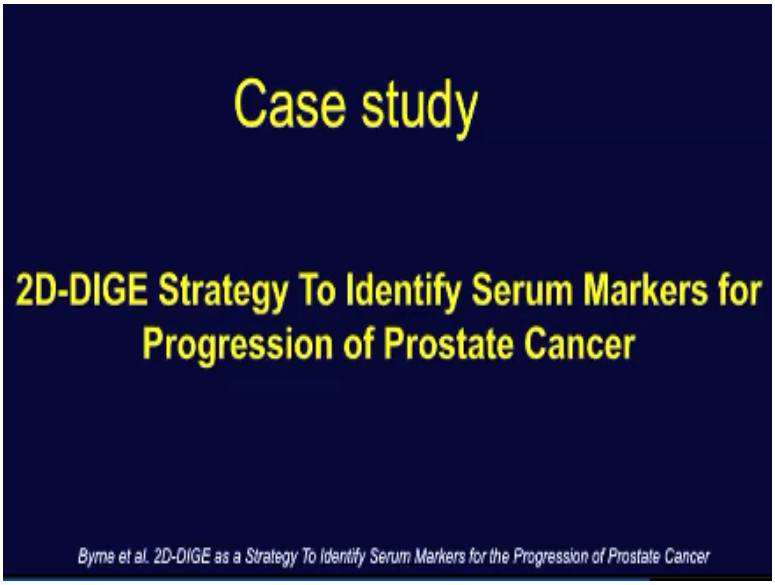
(Refer Slide Time: 24:29)



Once the spots are analyzed and exercise from the gel then the same tradition you have to follow you can use an of the mass spectrometry platform and then obtain the MS spectra for further analysis using different type of bio and pharmatic tools such as mass spot. So overall the DIGE method is far more superior in terms of the reproducibility as compared to the conventional two-dimensional electrophoresis and for the quantitative accuracy.

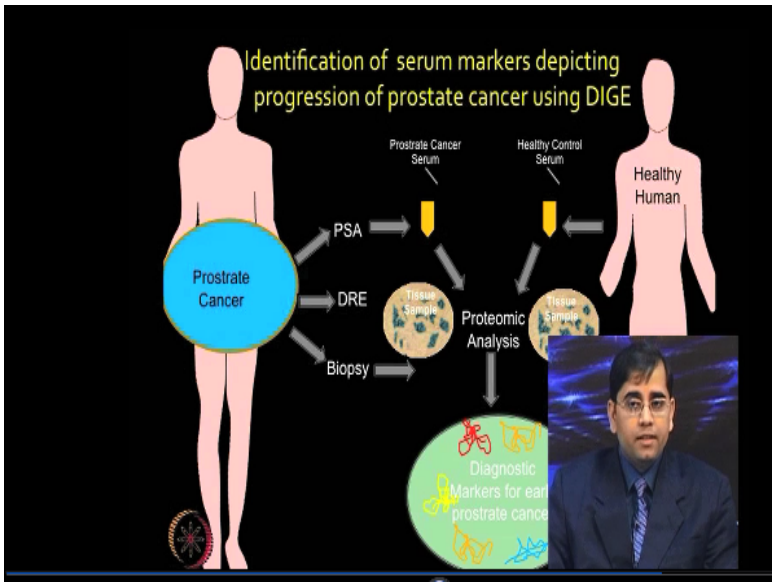
So therefore applications of 2D DIGE can be found in virtually all major biological research areas. If you see the recent publication you will appreciate there are several papers on each of the biological system for different, different type of applications whether self signaling or looking at developmental biology, looking plan proteomic analysis, neurosciences, clinical studies, different type of diseases including cancer you will find there are hundreds of publications available which have employed the power of 2D DIGE technique.

(Refer Slide Time: 26:09)



So let us talk now a new case study case study 3 on 2D DIGE as a strategy to identify serum markers for the progression of prostate cancer.

(Refer Slide Time: 26:31)

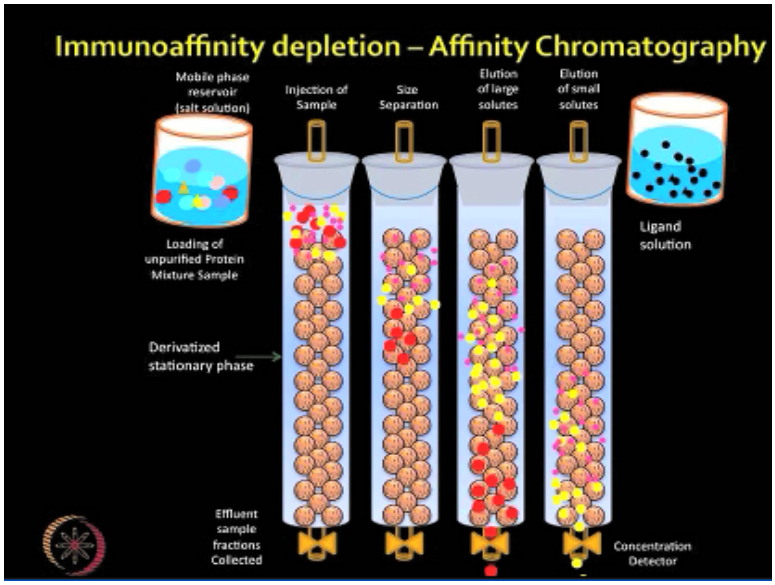


So this study by an Byme et al published in 2009. So in this study author aim for identification of serum markers by depicting the progression of prostate cancer by using difference in gel electrophoresis techniques. The prostate cancer is recognized as a significant problem in older male population. The prostate cancer screening rely heavily upon testing for the higher level of prostate specific antigen also known as PSA within the peripheral circulations.

So PSA is very sensitive marker, but there are lot of discussion on reliability and the specificity of PSA for the prostate cancer. Reason being that the level of PSA is also high in

benign prostatic hyperplasia or prostatitis. So therefore there are lot of discussion whether one should rely on only PSA for detection of the prostate cancer. So this study aims to identify some new markers in the prostate cancer by studying the serum proteome analysis.

(Refer Slide Time: 27:53)

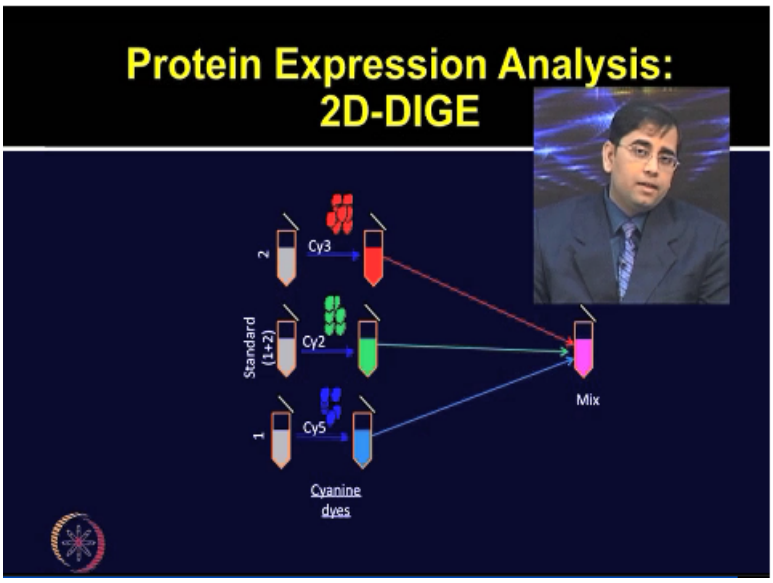


So as you are aware and in fact we have discussed the protein preparation from the serum earlier. So each of the biological samples posses lot of technical challenges and serum is one among them where presence of highly abundant proteins such as albumin and immunoglobulin they result into the masking of low abundant proteins. So to eliminate those high abundant proteins authors used multiple affinity removal system from the Agilent technology.

And they removed most highly abundant proteins from the serum sample including albumin, IGG, antitrypsin, IGA, Transferrin and Haptoglobin. After the abundant proteins were depleted from the serum sample then authors moves for the protein extraction and further analysis. So the differential proteomics analysis was performed in the 2 different cohorts of histologically confirmed prostate cancer with different grades of the disease.

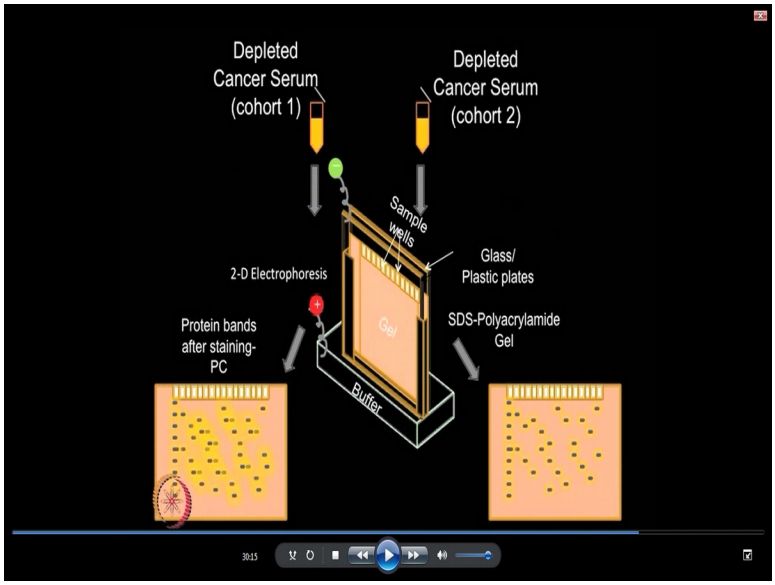
So they used the patient with 2 different grading system based on the lysine grading. So the Lysine grading system that is used to help and evaluate the prognosis of men with the prostate cancer. So depleted serum samples obtained from the patients with lysine score 5 and lysine score 7 were used for comparison and further analysis.

(Refer Slide Time: 29:47)



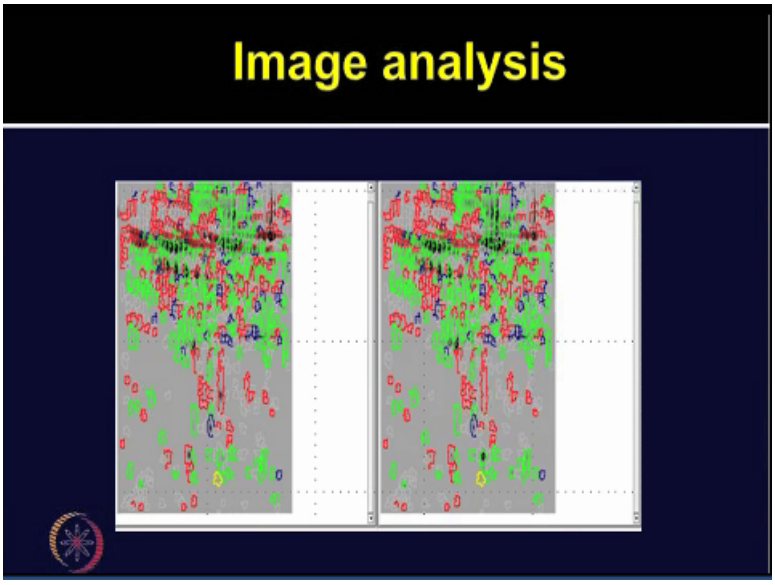
As you can see in the slide these samples were first labeled with the Cy3, Cy5 and also the internal reference pools were made which were labeled with the Cy2 dyes. These samples were then further mixed.

(Refer Slide Time: 30:08)



The depleted cancer serum from first cohort of lysine score 5 and second cohort of lysine score 7 those were mixed separated in the first dimension and followed by protein (()) (30:23) in the second dimension.

(Refer Slide Time: 30:28)

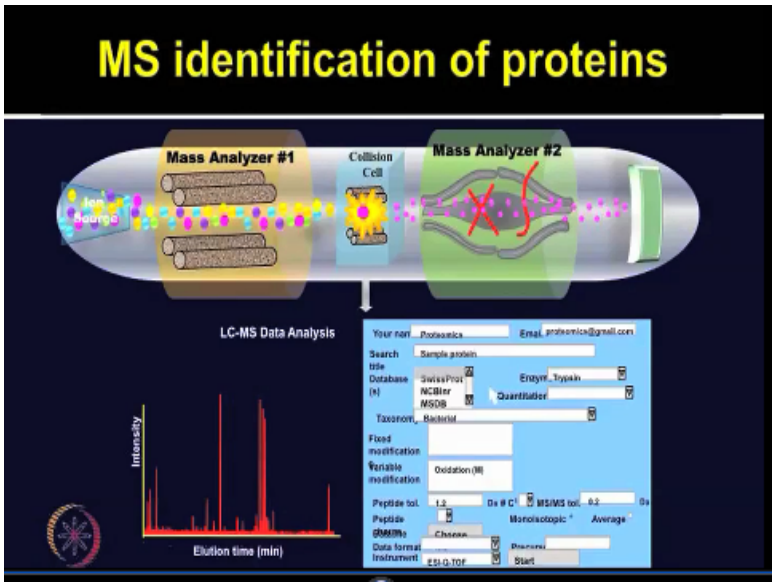


When authors analyzed these DIGE images they found that 63 protein spots were differentially expressed between the lysine score 5 and 7 cohorts and 13 of these proteins were statistically significant among these 2 population. So as you know analysis of these gels is always challenging especially if you are looking at the conventional 2D gel where you have separate gels obtained from each of these groups.

But analysis in the dye gel is more automated so if you remember our previous discussion in the dye gel analysis this analysis is more automated and more straight forward, but still we have to go through individual spots and you have to look for the how real, how significant those changes are and you have to look at the 3D view the whole spots to ensure that it will reproducible among various control and treatment groups.

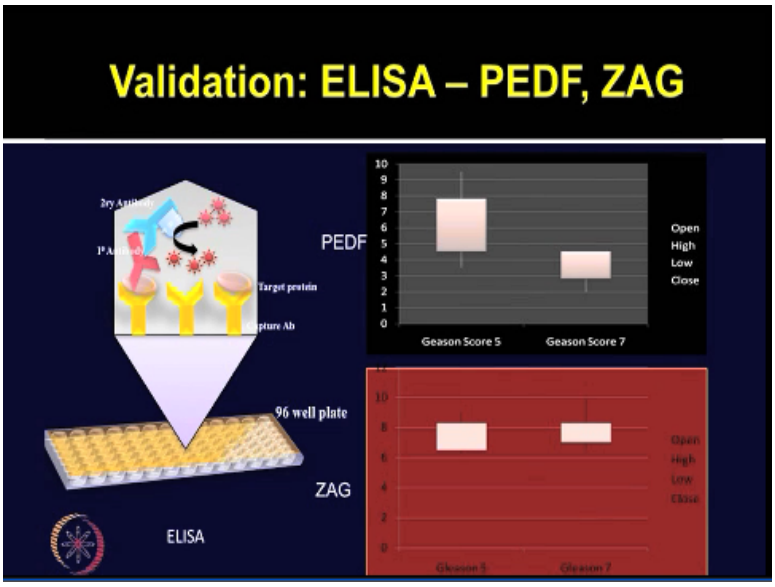
So that the different label of analysis is performed which we have talked earlier, but this just shows you the final output that 63 spots after all the analysis steps were considered significant.

(Refer Slide Time: 31:49)



After 2D DIGE gel image analysis authors exercise those spots and subject it for the mass spectrometry identification of proteins. So the proteins exercise from the gels were analyzed by using Finnigan LTQ mass spectrometer and data from these MS experiments were analyzed by using bio-works browser by using (()) (32:16) program.

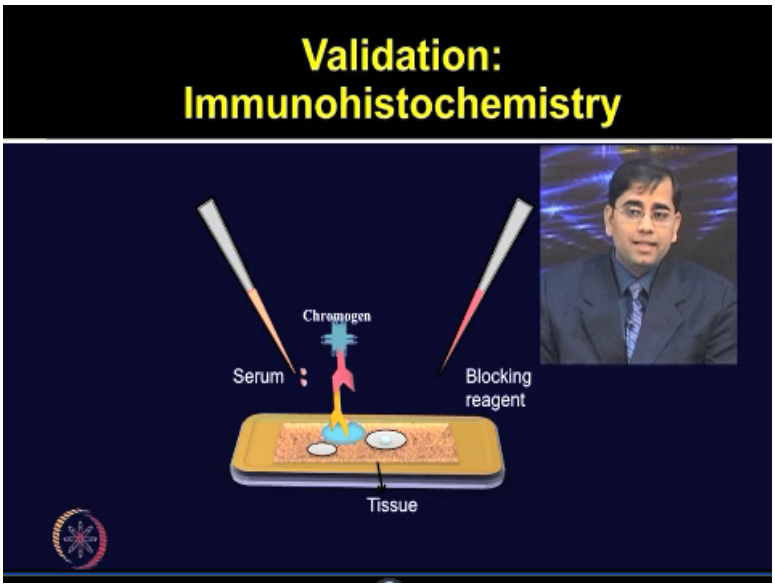
(Refer Slide Time: 32:18)



For validation authors employed various techniques including Western blots and Enzyme-linked immunosorbent assay or ELISA and also immunohistochemistry. So the pigment epithelium derived factor PEDF and Zinc alpha 2- glycoprotein also known as ZAG those proteins were further validated by ELISA technique. So the PEDF levels were quantified by using ELISA kit and results demonstrated as you can see in the slide that statistically significant decrease in the PEDF in the Gleason score 7 depleted serum group.

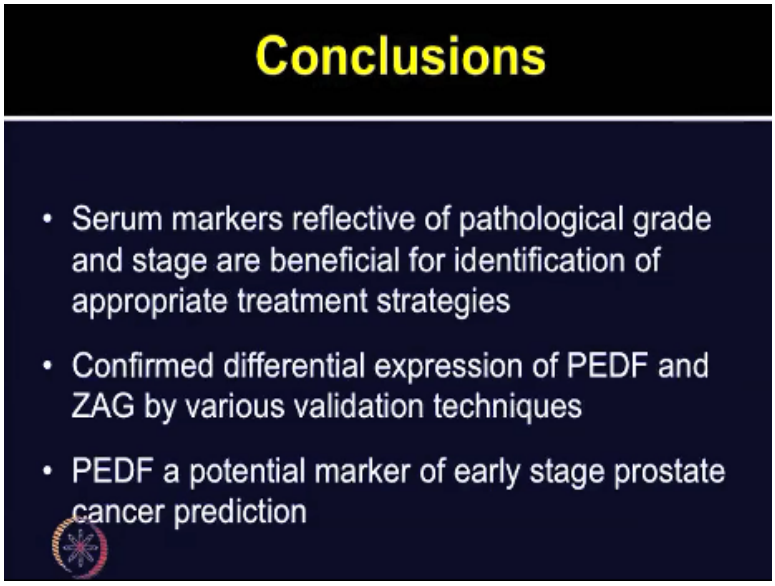
Whereas the results for zinc alpha-2-glycoprotein ELISA which is shown in the red in the bottom panel that indicated 1.44 increase in the zinc alpha 2 glycoprotein absorbance in the Gleason score 7 group. So these studies this ELISA validation confirm their findings from the 2D DIGE experiments.

(Refer Slide Time: 33:40)



Authors also employed Immunohistochemistry or IHC for validating the pigment epithelium derived factor PEDF and zinc alpha 2 glycoprotein so that they are very confident that the protein which they have identified from the proteomic profiling those are real and those are tested on the independent tissue samples.

(Refer Slide Time: 34:12)



So from this paper the major conclusions were that serum markers which are reflective of the pathological grade and stage could be beneficial for the identification of appropriate

treatment strategies. Authors confirm that differential expression of PEDF and ZAG can be performed by using various techniques such western blots, ELSA and Immunohistochemistry based on the STD and the follow up experiment they concluded that PEDF could be a potential marker of early stage prostate cancer prediction.

However, more studies and follow up required on the large number of patients before it can be established as a good biomarker. You may appreciate that there are lot of power of these techniques and these can be employed for any biological application you pick an application of your choice and I am sure you will be able to answer those by employing 2D DIGE techniques. Thank you.