HANDOUT

LECTURE-15

TWO-DIMENSIONAL ELECTROPHORESIS: IMAGE PROCESSING AND DATA ANALYSIS

Slide 1:

This lecture will cover two-dimensional electrophoresis, image processing and data analysis.

Slide 2:

The previous lectured have covered the following:

- Gel Based proteomics: The following topics have been discussed- SDS-PAGE,
 Blue-Native PAGE, OFFGEL fractionation as well as 2-DE.
- 2-DE workflow: Procedures covered included the steps of a 2-DE experiments including various details of performing protein extraction, different types of challenges associated with the sample preparation followed by isoelectric focusing, equilibration and SDS-PAGE running and analysis, followed by gel staining, visualization and spot analysis to obtain meaningful data.

Slide 3:

Lecture Overview

- This lecture will continue to focus on 2-DE.
- Image processing and data analysis step-by-step and various types of challenges associated with image processing will be covered.

Slide 4:

- This example is a gel run (of a human serum sample) when the experimental conditions have been optimal.
- Such a run will yield an image as seen above, with well resolved spots throughout the gel.

- But many times, experimental conditions are not this well adjusted. Many things can
 go wrong and cause artefacts, for example, interference from salts, nucleic acid
 contamination etc. Various types of vertical as well horizontal streaking can be
 observed.
- These issues, while small, can interfere with optimum gel analysis during image processing.
- Furthermore, image processing is not so tedious when one has a small number of gels to analyse. But often, a large number of control, and treatment gels have to be run in a single day, making consistency hard to maintain for various parameters, including, say, the point that the dye front migrates to or the protein resolution.
- Such a lack of consistency, however, can be partly corrected using various image processing tools in the software programs available. The following sections will talk about different types of image processing tools, how to prepare the images for data analysis and then how some of the commercially available software can be used for obtaining meaningful information (such as a fold change in protein expression, or the presence of unique or modified proteins in one sample group, etc).
- As previously touched upon, these software also allow statistical analysis of proteins of interest, allowing the identification of spots that can be further analysed through mass spectrometry. All of this makes data analysis far more time consuming and technically demanding than the gel run itself.

Slide 5:

- This lecture contains the summary of the discussion about image processing and data analysis using a commercial software program, the 'Image Platinum' software.
- Image Processing using this software helps address the issues that arise while running gels.

Slide 6:

 During the course of this discussion, the stepwise procedure of how to take the scanned images and the raw tiff files and then align those, followed by spot

detection and various types of manual refinements while processing the images and finally, obtaining meaningful information using various statistical analysis will be covered

Discussion with Dr. Shrinivas, Application Specialist from GE Healthcare.

ImageMaster platinum.

- This software plays a key role in the analysis of gels, especially so when we have multiple numbers of gels, it is impossible to analyse manually.
- Image Master is a flexible solution for the comprehensive visualization, exploration and analysis of 2-D gel data.
- Two modules are available in this software. The first one is ImageMaster 2D platinum and within this we have another module ImageMaster platinum which is enabled for the DIGE application also.

Introduction

ImageMaster is a flexible solution for the comprehensive visualization, exploration and analysis of 2D gel data

Two modules:

- either exclusively for conventional 2D gels
- or for the use with DIGE gels and conventional 2D gels



 Note: If you want to full-fledged DIGE analysis one should have DeCyder. For conventional 2-DE, ImageMaster platinum (IMP) is a better tool for analysing the gels.

An example is provided wherein the IMP7 software is used to analyze 12 gels.



- A. The various types of tools one can use, how the images look like and what kind of parameters are there on display.
- Editing: This is very important because during scanning one may have some extra
 portion of the gel as well as the streaking of the gels at the cathodic end or the
 anodic end. These all should be removed from the gel.

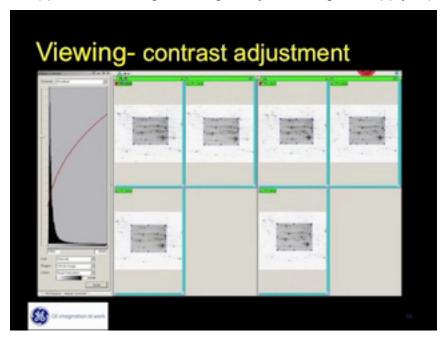
Analysis Workflow

- Add images
- Edit images (crop, rotate, etc.)
- View Images
- Construct match hierarchy
- Process images (detect spots, landmark, match)
- Define classes
- Data analysis
- Export reports, pick lists, etc.

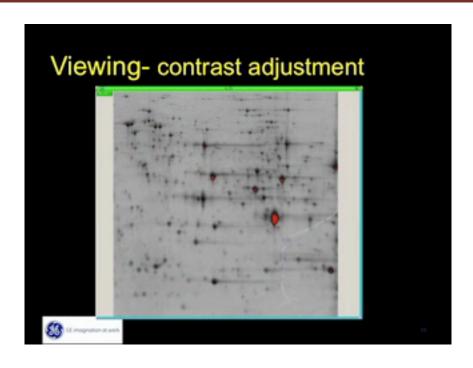


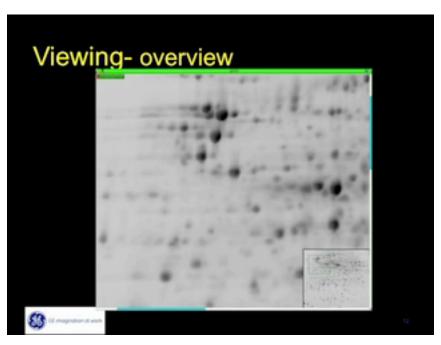
- Editing is very important, then visualization of the image.
- Construct matches hierarchically. There is a procedure to create a match set and how to create the class set. There are two different processes.
- Then process gel images where we can detect those spots.
- Define Classes
- This is followed by data analysis.
- Note: Preparing images for analysis is very crucial because, while all the software claim that this process can be entirely automated, they still do not filter out noise such as dust particles, which can be detected as small spots on the gel. So one needs to have both automated and manual validation and balance of each to ensure whatever spots are picked up are the real spots. Adding the Image.
- In the image analysis workflow, first we will go through add image, how can one add the images to this software. There is an option "image pool" as you can see here. From there you can add any image stored in any area of your computer. The software can analyse .gel, .mel, .tiff kind of images. HoweverExcept simple jpg or compressed tiff. These kinds of gels can't be edited. It requires high resolution images. The image resolution should be atleast 16 bit.

- While adding your images it will ask you what stain you used for your gel. So we
 need to mention the colour especially. Within the software we incorporated different
 stains like coomassie, silver or different fluorescent colours which you have to
 mention. Once you mention that then you can open your images.
- By assigning a particular staining, one can assign pseudocolours for these images.
 So it helps viewing the images. Adjust the contrast: Under the 'view' list of options, there is an option to change complete contrast and brightness.. The change made can be applied to all the gels being analysed using the "apply" option.



 Saturation: This is indicated by the red colour, as seen below Areas of saturation should be viewed using the 'zoom' option, to determine if it is a single or multiple spots clumped together.





A. "Profile view". This c provides the profile or the 3D view of a single spot. This option is useful in identification of the dust particle contaminants.

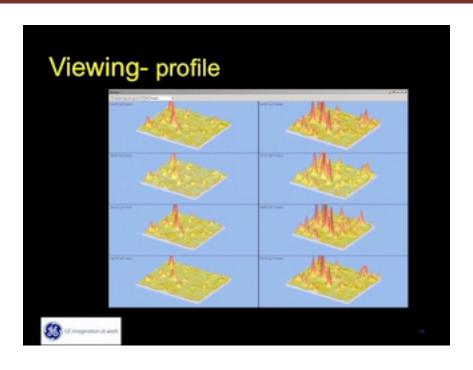
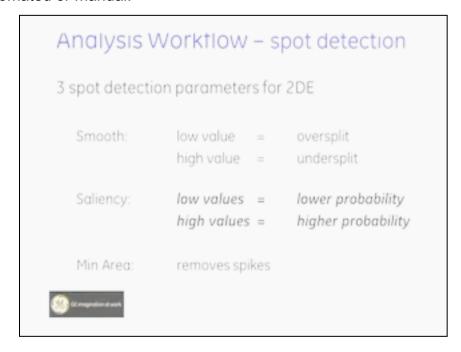
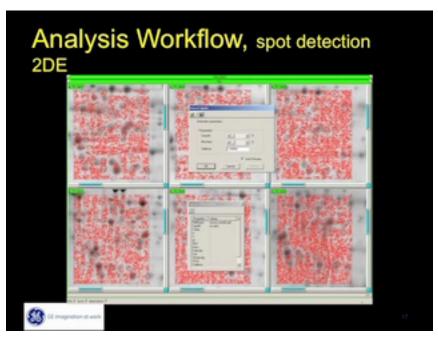


Image processing:

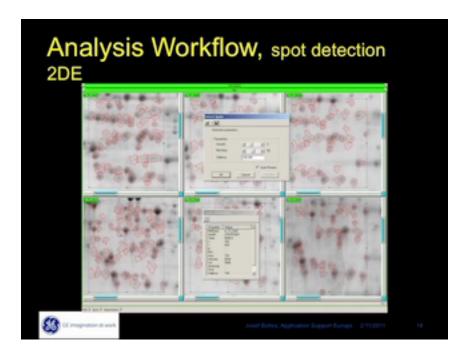
- This involves spot detection, matching and quantification.
- Can be automated or manual.



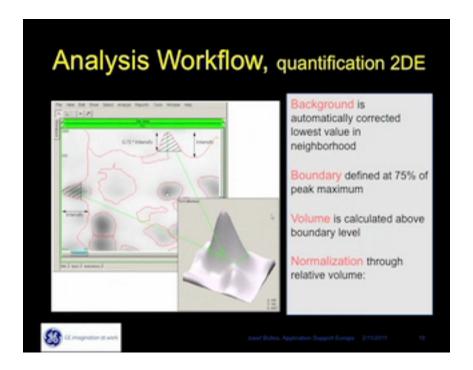
- Automated detection: This involves defining parameters smoothness, saliency and minimum area Smoothness can be set between 2-3
- However, assignment of the saliency value is a bit more tricky. Saliency is a very critical parameter. It is the parameter that distinguishes foreground from background.
 - A. Minimum area can easily remove the dust particles, because any protein having minimum area above 5 atleast. When the minimum area is assigned as 5, this removes dust particles from consideration.
- Shown below is the results of an automated spot detection wherein the parameters
 were not ideally assigned. The red boundaries denoting the spots are software
 assigned and you can tell at a glance that spot detection has not been correctly
 performed.



 After that, parameters were modified and manually assigned and now you can see that spots are perfectly detected and perfectly boundaried.



- Note: The software should be used to first detect spots, and whatever is missed out can be manually detected.
- Another option "exclude" allows one to exclude any particular area in analysis.
- It means you are interested in certain region of the gel then one can include only that region of the gel.
- In this experiment may be they are interested in this particular area only that's why
 they have included this much data.
- This is the 3D view, we can clearly see that the background is automatically removed here because whatever boundary we are putting which you see in the 3D image it is not taking the bottom of the 3D. That is the protein peak. It is taking some portion and leaving some portion and it is creating a particular boundary. So that the total volume it is going to take is above the boundary only. It does not take below the boundary so that we are eliminating the whole background.



Spot quantification

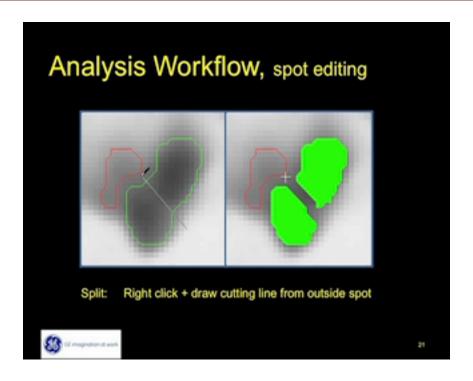
The software can assign intensity, area, slope and volume ratio to each spot based on algorithms.

- Background is automatically corrected lowest value in the neighbourhood.
- Then boundary; it is defined at 75% of the peak maximum. So one can easily remove the background.
- Then the volume is calculated above boundary level only where you can see the green boundary above this only we are calculating as volume.
- Then normalization through relative volume that means the relative volume from all spots this 25% is reducing. So remaining whatever volume is relative volume.

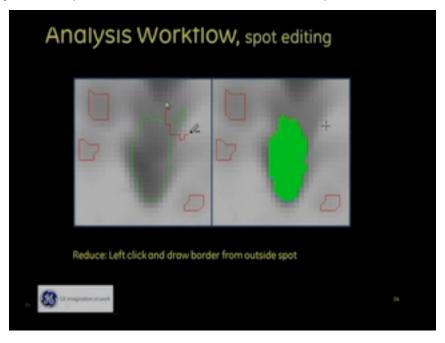
Spot Editing

- Most of the time if it is possible, avoid spot-editing.
- Use alternatives like composite spots by defining multiple matches and propagate spots from one match set to another.

- Like once you are matching, these are all not matching perfectly and the same gel
 you can try to match another time and you can introduce few landmarks so that it is
 easily matched and detected.
- So, your recommendation is that one needs to rely on the automated analysis and because the numbers of gels are more it is difficult to do the quality control checks with the manual way.
- By defining various landmarks, by repeating the process several times one can increase the precision of analysis.
- Then if you would like to edit your spot how to edit the spot first. First of all this the way we can split the spot.
- There is a tool for spot splitting and then you just draw where ever you want to split then it can be created as two spots now.
- In this way you can edit your spot and as well as lets say you have to do many spots, by this way you can draw your line from both sets so that it can be much together.
- Another option in spot editing, if you are not satisfied with your boundary and you
 want to enlarge your boundary, you do not need to delete and re-draw your
 boundary. Instead you can just mention how much area you want to extend to this
 spot so that you can expand the volume of your spot.



 In the same way if you want to decrease some portion of your spot like as you can see there is a tail sort of thing but you think this may not be your protein portion then one can easily cut that portion and hence can reduce the spot size.



- So, defining the accurate spot boundary is crucial and different types of software views like 3D views and profile views can actually help one to take that decision of what is the best boundary one can define for the spot.
- Matching: Once you have finished your editing process, now this is the time to match whole gels so we can place one or several landmarks to support automatic matching. But if you put landmarks, it can be very useful for matching. I will show you images how it effects the matching, then keep landmarks at minimum. If it is possible try to minimize it to 2, 3 or 4. Otherwise it will create some discrepancy in your spot searching.

Analysis Workflow, matching

Place one or several landmarks to support automatic matching.

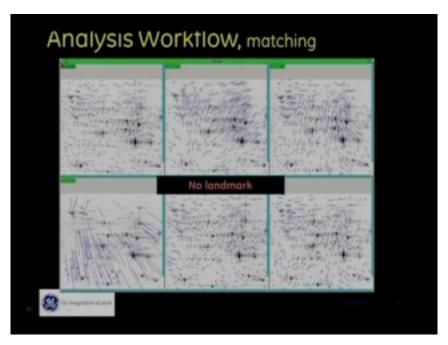
Keep landmarks at minimum.

Place on well defined smaller spots wich are crearly corresponding.

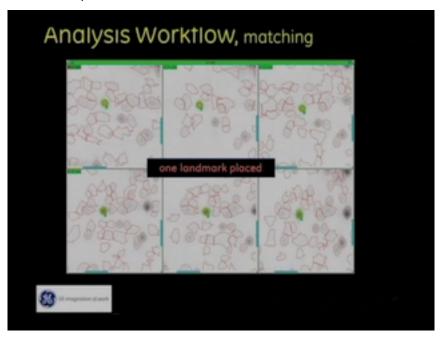


- So I think it will be helpful to have landmarks in various regions of the gel so that one
 can actually align all the gels by using these landmarks. They should be placed on
 well-defined smaller spots which are clearly corresponding.
- Once you match these kinds of vectors you can see. What does it mean? This is
 deviating from the reference gel to remaining gels, how it is deviating for that
 particular protein whether it is one side deviating or it is differently deviating you can

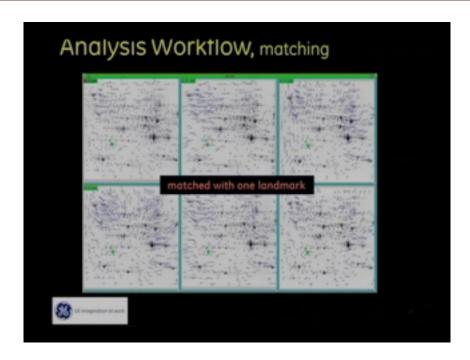
see here. It will show you overall trend of the matching. So there is no landmark at all.



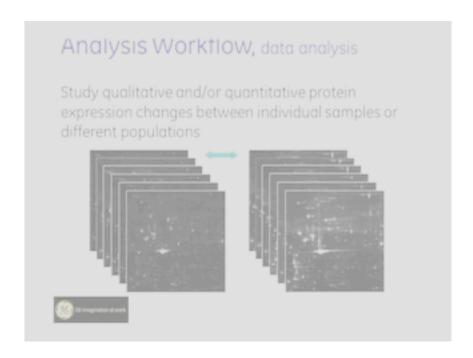
 You can see once we put the landmark how drastically things changed. There is no tail. This is how it helps.



So, this is another thing matched with one landmark.

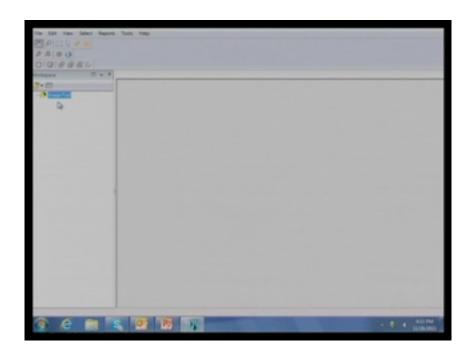


- So the vectors look well organised and have similar orientation as compared to the ones we have seen earlier where there was no landmark.
- After matching everything, how to design your experiment? We will next cover how to create a match set.
- From our image pool we have added images and after that we have to create match set, we have to create our class. We will talk about it in detail in our software demonstration. These are all the analysis workflow, the complete data analysis, once you get the complete data from that what are the flow fold regulations, what are the statistical parameters, then you can count all the parameters and you can select your protein of interest.

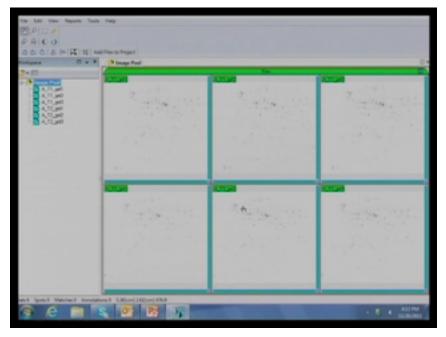


ImageMaster platinum demonstration

- This is the software layout.
- Here there is an option like image loader.
- Go to the place where your gels are stored and add the gels. As I have multiple gels
 here, I will add few gels to the software to analyse.
- I am processing 6 gels at a time. These are all 6 gels basically replications of 2 gels.
 3 replications for each gel. One is control and another is treated. As we discussed in our slides this is particular thing asking for staining. Here we can give coomassie, silver etc.
- So as it is stained with silver, I will opt silver and apply the same colour for all the gels.

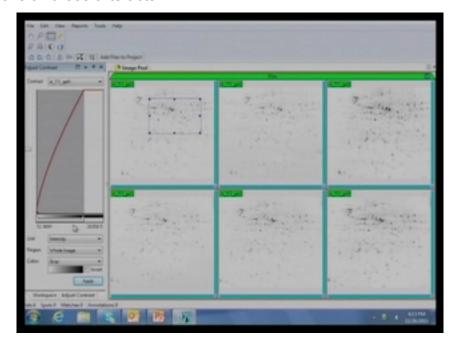


 Now you can see the 6 gels at a time in this overview. You can't see proper spots in all the images. They are very bright images. Now we need to edit the images for visualization. We need to change contrast, brightness. Now more spots are seen.



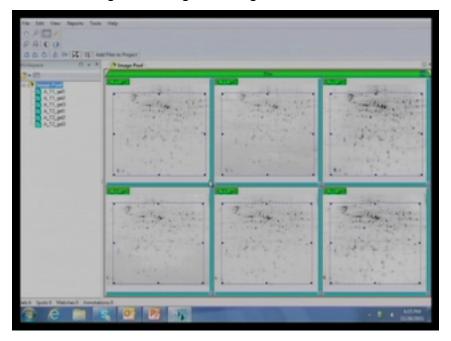
 Otherwise there is another option. You can select a particular area and then change the parameters.

- Now you can see real time changes in the particular spots. After you are satisfied
 with the parameters you can click apply so that it can be applied for all gels.
- So now we can see more spots clearly on all the images. Now we should not disturb
 in our original images so we need to make copy of these particular images.
- For that we can crop and create new images. As we have some artefacts in our gels
 we can remove all those artefacts.

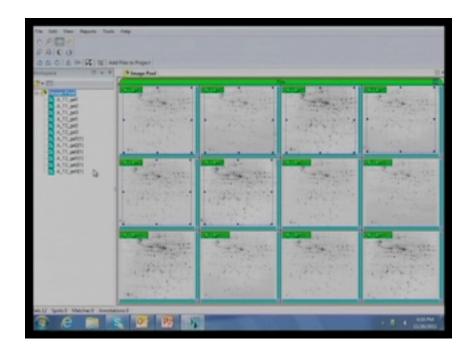


- So there are some gel migration issues and some gels do not run uniformly you can crop the images and you can store a copy of the fresh image without changing your original image.
- Otherwise once you are doing your analysis you have changed your original image you may lose your raw data. This should not happen and that is why we are going to crop them.
- When you are selecting in a particular gel it is going to be selected in a single gel.
 Instead of doing single selection just press on shift+control and select area which
 can be common in all gels. So that it can be common in all gels so that you can
 select at a time in all gels. Then you can adjust them in the same way. After pressing

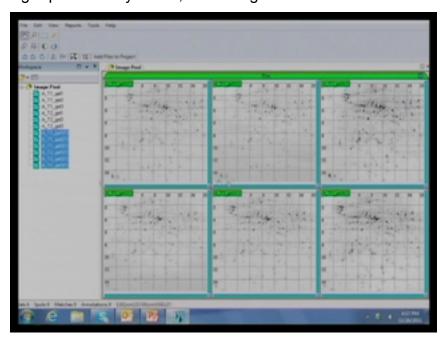
shift+control you can edit all boundaries at a time. Otherwise if your editing a single boundary at a time there might be image-to-image variations.



- So, manual adjustment is not required because we can do this automated way so that all the processing will be uniform.
- Then we can crop them using the crop option. And click yes to create new images. It
 a very quick process. So now we can see all edited images here with extension of
 (1).



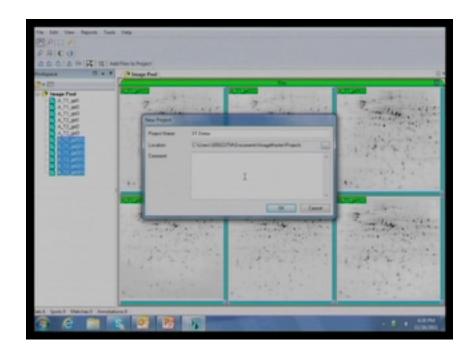
- The original images have been stored and created new images with the crop.
- After that we close all the images and then the edited images are opened. One can
 create a grid lay on the gel from view option so that one can judge whether all the
 gels are showing reproducibility or not, in which grid I have what number of spots.



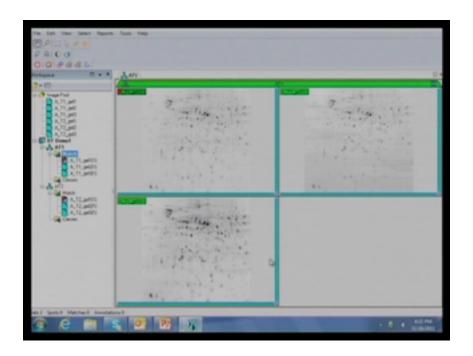
- Also helps to decide which landmarks to use, also which area one needs to look around.
- This is very helpful tool again. If you want to remove the grid overlay click the same option. Another thing we can talk about is profile. One can click the profile and wherever the mouse goes it shows the particular spot profile, how it is looking. Initially we can see the gel quality. Then you can remove that particular tool.



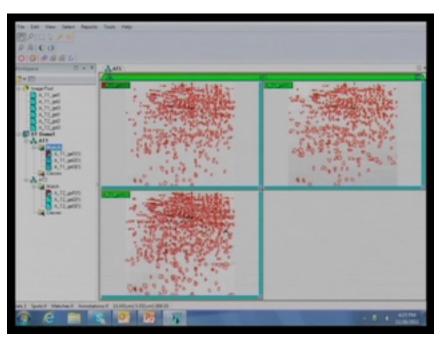
- Now we are going to plan the design of our experiment and how we can manage our gels. From the project we have to create new project. The project here we are doing is the IIT demo1.
- If you have any description about your own experiment, if you have specific details
 about the project then enter here and can be stored in this project. So now in this IIT
 Demo1 one can create the match set here by clicking match set. You have control
 and treated.



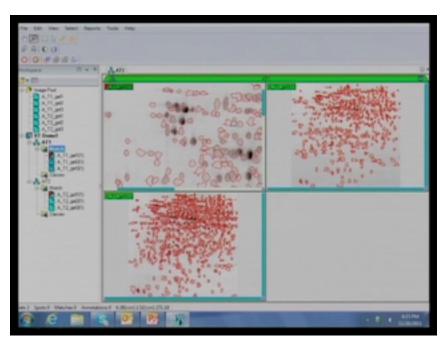
- So, you have control and its 3 replicates and treatment and its 3 replicates.
- Then create another match set. Two folders are created and shift the corresponding gels to these folders. Select the corresponding images and just drag them into the folders. Here you see one gel is half red and half green in one folder and it is the same in the other folder. This means this gel is set as reference gel. The remaining 2 gels are going to be matched against these two gels.
- Any gel can be used as reference gel. There is an option in view where you can change.
- Now display them and you have all of your gels. Once you have the gels we can directly detect the spots. Now the 3 parameters smoothness, saliency and minimum area. If you increase the smoothness it is going to under-splitting. If decreasing the smoothness it is going to over-split. So it is inversely proportional. Its average value is from 1 to 5 but 2-3 is sufficient to start. And then saliency; one can set upto 100-150 but the good value to start is almost all 15-30.
- Initially I am just giving 15 then minimum area 5 is good enough. If you think in your spot there are more artefacts you may increase this also.



 Now all spots are detected. If you are satisfied with this particular detection then its fine.



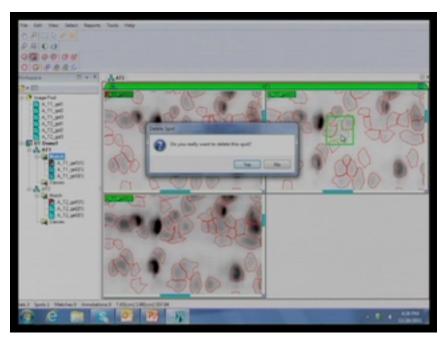
 Otherwise you can zoom this particular region and see how the spots are boundaried as well as how the detection is there.



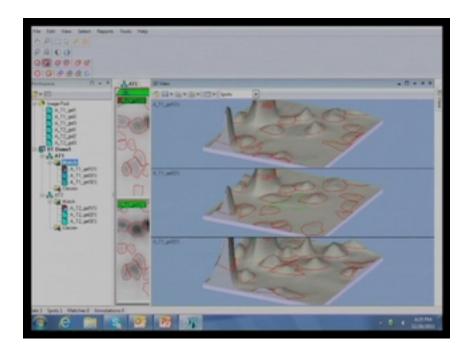
- A. Can you apply the same zooming parameter for all 3 gels?
- B. Yes, definitely. First come back to original and select a region and the same region is zoomed in all the gels.



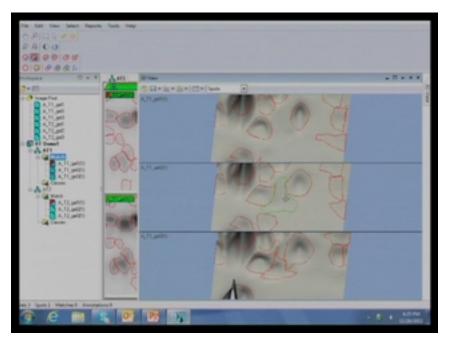
- I think it's very useful to see this way. Now the parameters are well defined because
 it has properly defined boundaries and there are some spots which on can look
 around.
- Actually you want to edit these particular spots as you can see that there is no protein.
- But it is showing as a protein one can delete these kinds of spots also. One has to go edit the spot. There is an option "edit enable". You can go there and delete the spot you want to delete.



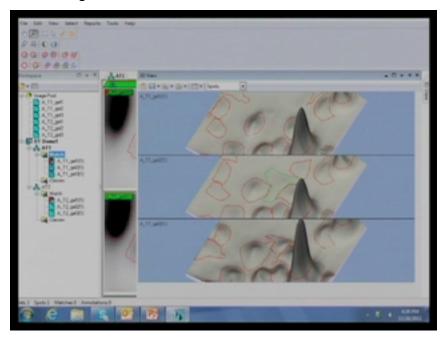
- But before we take that decision can we have a look at the 3D profile to ensure that we are not deleting the real spot.
- · Click on 3D view.



- So the spot which we are interested in discussing is highlighted in the green boundary.
- But it is only present in this particular spot. It is not there in the first image but we
 can see some portion of this spot in the 3rd image. That is why it is detected there
 also.



- These are 3 replicates of the same protein and hence they should be same. But somehow they are not and so we can delete these spots.
- When we finally see the statistical data where it is there in both gels as well as treatment also so that one can go with the statistical parameters. This will be helpful.
 So need to delete also in this case.
- Otherwise I can zoom more so that one can see it. Now this is the intense spots which you are visualising.



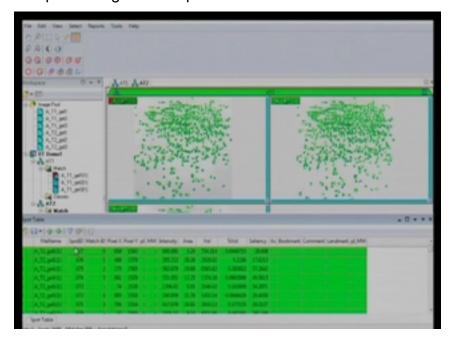
- Actually we can zoom out and go back to the first gel. So lets say in a region we
 have missed out some spots, like software missed out then there is an option to add
 the spot as well. Let me select a particular area and let me zoom that. Now you can
 see that particular area in all gels.
- If you want to include a spot wherever you interested you can add. Like this end of extra portion can be deleted easily.
- Now we find all spot boundaries are ok and every thing is fine.
- Now we can match these gel and check for the reproducibility of the 3 gels. Zoom
 out to the original position and match them using the match option.

- We are matching control gels only to the reproducibility with in a group. Suppose you have 5 or 6 replications instead of 3 replications, out of 5 or replications one can remove that particular gel and they can take the remaining gels.
- Now in these two gels there are few vectors, which you can see. These you can remove by adding the landmarks.
- Now this is the landmark option. Initially you have add landmark in your reference gel only. After adding the landmark, you have to adjust the landmark to the same position in the remaining gels.
- Again match them. You will see that from 146 it has increased to 746 by adding a landmark and our vectors are matching well.



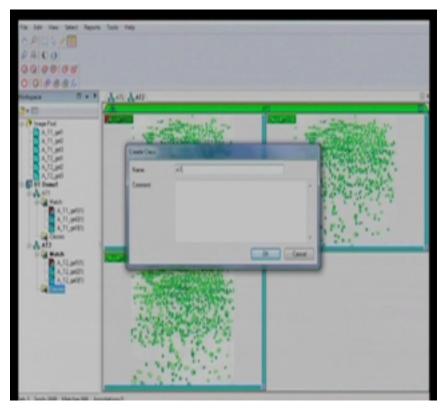
- You can see there are few vectors but they are still there. These are all moving to the same side. So that means there is no irregularity. One needs to go through each region carefully.
- So if you can add one more landmark here it can be removed very easily. This is what we can do.
- And now we can save this thing and apply the same parameters to our treatment gel also. If you can detect your spots automatically it gives same parameters earlier which you used.

- Similarly you can go to each gel by zooming, 3D view and selecting the spots and non-spots and directly match them.
- 1217 matches are found. These gels are more reproducible as compared to the other set. There are no vectors even with out giving landmarks. So this is very fine gel.
- The replication with which we did was confirmed. These replications are ok. Now one can analyse in the class analysis between these two classes like interclass and intraclass.
- Now I am creating the classes here. Before that I would like to go to the report that is
 the gel table. These are all the gels which are present here. By pressing shift+A you
 can select all the spots and go to the spot table.

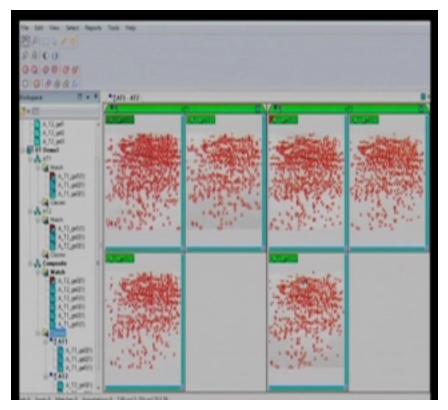


- The parameters are described here in this spot table include the file name, spot Id, pixels, match Id, x-axis, pixel at y-axis, pl, molecular weight, intensity of particular spot and area of that particular spot and volume of that particular spot, percentage volume of that particular spot, saliency of the spots.
- If you are satisfied with these values you can go ahead for next level of analysis that is class analysis.

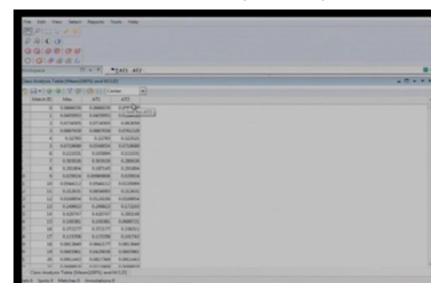
Now I am going to create classes named 81 and 82 which are control and treated.
 And shift the gels to the corresponding folders. Another match set I am creating so where we can move easily before going to class. So I am moving all 6 gels into composite match set and now create again the same way, classes.



• Now I am shifting all the gels to the corresponding folders. This is for class analysis. We can see all the gels together, 3 gels from control and 3 gels from treated. No it is taking values as single gel of 81 and single gel of 82. It will average the first 3 gels and then average the remaining 3 gels. So a single value is obtained from the control and a single value is obtained from the treated. And quickly match them again. One can remove the vectors by adding landmarks, same process as we did earlier.

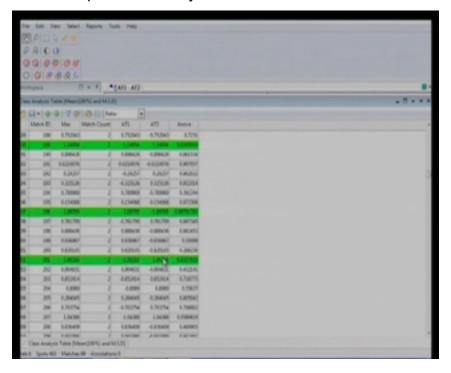


• So go to reports where you can see the analysis classes and then go to table. Now we have the match Id. Match Id is a particular spot id as well as the maximum area.



 So, this software gives you option for various statistical analysis. You can also see the annova value. Basing on this we can easily select the proteins of our interest as

well as you have the fold-change also. We have 1 fold up-regulated and 1.2 down-regulated. You don't have to go to each spot. You can filter them easily. We are going to take maximum annova value of 0.05 as statistical significant value and so those min-spots can be highlighted. All the spots with this particular annova value are highlighted. This table can be exported to excel and from there you can have your data. This is the output of the analysis tool.



Slide 7: Summary

The discussion on image processing using Image platinum software was useful for you to understand the details required for image processing and how various types of parameters one need to pay attention during the analysis. So in today's lecture in summary we discussed about the challenges for the image processing associated with 2-DE, how one can process the images efficiently, how one can analyze these images stepwise using various software. Most of the software work with the similar principle and to give you a feel for data analysis we showed you a software demonstration during the discussion with Dr. Srinivas from the GE Healthcare by using ImageMaster platinum

IMP7. At the end you must appreciate that one needs to do different types of statistical analysis to ensure that the identified spots can be further pursued for research and those are going to be further processed by mass spectrometry and if they can provide some meaningful information then one need to have full confidence and that's why we need to do statistical analysis.

Slide 8:

So I would like to acknowledge Dr. Srinivas from GE Healthcare for discussion on image processing and data analysis as well as for giving us a demonstration of software for 2-D gel analysis.

Slide 9:

In the next lecture we will continue our theme on gel-based proteomics and I will introduce you with a new technology on 2-D difference in gel electrophoresis or 2D DIGE and then we will continue our lectures on 2D DIGE and its image analysis.