LECTURE-31

MICROARRAY WORK-FLOW: IMAGE SACNNING AND DATA PROCESSING

TRANSCRIPT

Welcome to the proteomics course. In today's lecture we will talk about microarray work-flow the image scanning and processing. This is in continuation of our discussion on microarray work-flow.

A - Prof. Sanjeeva Srivastava   B- Mr. Pankaj Khanna

A - This is my pleasure to introduce Mr. Pankaj Khanna, manager-application support from Spinco biotech private limited. Today we will be talking about GenePix microarray scanner which is from molecular devices and Spinco is distributor for the same. So we will have a discussion about microarray scanner as well as the software how to control the hardware. so we will talk about various basic concepts as well as a live demo of software interface so that you get a better feeling about how this whole process is controlled and performed so Pankaj welcome to this discussion about microarray scanners and software.

B - Thank you Dr. Srivastava, it's our pleasure. At Spinco biotech we have molecular devices as distributors for last many years and we are dealing with the scanners for the many years at the level of microarrays as you have already described that it is used in the form of microarray. We are dealing with the image acquisition to the level of data analysis for the microarray applications so let's go through little bit on what is GenePix actually can do.

A - It will be useful if you can just first give some steps for the DNA microarray experiments although it is proteomics course and I have been more on the protein microarray and things but this whole technology got initiated from the DNA microarrays so just so that student have some idea about how the DNA microarray perform just give

some brief steps about overall DNA microarray technology and also like how and what GenePix can help in the microarray process.

B - Sure. So basically the DNA microarrays steps are have been divided into six basic steps starting from the manufacturing of the array so based on the slide and biology, manufacturing of the slide so if you want to go for the RNA expression chip you have RNA expression chip made, so if you are using a protein expression chip you are using a protein expression chip. So after the design has been decided next step is experimental design next step scientists think based on the biological question they are asking so based on that they choose which reference and which control to be studied, once you are ready next thing comes is sample collection and sample preparation which is then can be hybridized to pre-planned chip which is now ready to go for scanning in the form of image scanners so the four step involve actually the harder beginning where we vendors from the molecular devices come in place and this data once scanned in the form of raw data, now can be made data building, filtering as well as normalization followed by the last step of bioinformatics and biostatistics which help in the data analysis. So in this fashion basic these six steps are important for any microarray experiment.

A - So just to clarify I guess it's the content which is different whether it's DNA or protein but once we have chips in hand then the entire assay and the whole procedure is very simple. So manufacturing of the chips will be different for each of the content type, then experimental design one has to choose depending on the content and biological question what they ask and once these things are in place then once has to choose what type of biological sample, clinical sample they want to probe and then followed by hybridization steps. And then once experiment is performed then one need to acquire the images and one need to add hardware and scanners come into play followed by we need to data the data analysis and statistical part comes into play to identify the meaningful information from these experiments. So these scanners are user friendly and even the cost wise and overall performance wise they are quite good. They are used worldwide and I had been using for my research at different places and I am also

continuous user of this scanner. I am quite familiar with the overall configuration and the hardware parts for the GenePix scanner.

B - As you are using 4000B which is actually as you said the world's best hardware in a form of Cy3 and Cy5 dyes are being used, apart from that people use for different applications as well.

A - So can we talk little bit about the platforms which are used for acquisition and analysis of these microarray data?

B - So in the form of hardware as we have talked there are number of different possibilities in the form of GenePix 4000B to GenePix 4300 and 4400 these are actually doing the hardware part which allows the lasers to scan the chip so once we are ready with the chip so it will be interface between the CPU and the system which is allowing the data acquisition to happen and GenePix is being attached with a software called GenePix pro software which helps in understanding how the data is being acquired with the hardware usage. So basically another software which is also used is Acuity which is actually a third level statistical analysis software so in brief hardware is being attached with the data acquiring software called GenePix pro and this software now give the results to the acuity which further analyze for statistical analysis.

A - Yes. Obviously, you need one software which can help in the performing the scanning process and once the images are required then you need to obtain some meaningful information from those images and that is where I think software like acuity helps.

B - Yes.

A - So which technology is better for microarray and what do you think of unique features of using this hardware?

B - So GenePix microarray as such is based on two special designs one of them is non-confocal design of hardware and second is inverted chemistry?

A - So what is this non-confocal design?

B - Non-confocal design is like where you have spots in the same plane and the scanning is done called is as the 2-D. when you go for a tissue arrays or analysis where it is very higher so in the form of spot size especially at the level of imaging so there at the level of microscopy people use a technology of confocal spectroscopy so that they focus the laser at different wavelengths so at the end of the day they build a 3-D image but usually for all microarrays which is the highest is the tissue microarrays it is shown that all the slides in the same plane which is why it is confocal chemistry which resides in the same plane of the site. So that is why we choose non-confocal chemistry attached with the inverted chemistry of scanning which helps in the best results the best results in the form of signal to noise ratio.

A - So what are some of advantages of this design of GenePix stage?

B - So as we discuss non-confocal coupled with the inverted scanning how does it help is usually we see in the glass slide there are lot of small deformity which our eyes can't see so what we do is invert the slide so that it goes to the level of same scanning and they are supported at the level of the edges with the controllers which see the deformities at the glasses. It help in looking at the deformities and directly allows scanning to happen on the phase of scanning area so what happens is because confocal design attached with this inverted chemistry we get best signal to noise ratio so that your data is more valid for any analysis.

A - So this process can actually resist the any misalignment which may happen due to the design of the slide and scanning procedure.

B - Very true.

A - So different type of slides, which are compatible for these types of experiments?

B - As you have already said these systems are used worldwide and in different academic institutes people use various academic as well as commercial vendors so

ranging from the different applications to different vendors they support to name few for example Agilent, nimbolgen, corning etc. these slides are all compatible with them.

A - So I think it's very useful to have a platform where one can use the slide from different commercial resources. Otherwise, for example if one want to use the same scanner for doing DNA arrays and protein arrays many times it becomes difficult but as you mentioned like we can also use protein arrays from Invitrogen as well as DNA arrays from various commercial places. So I think it just shows that this platform is quite robust for wide applications.

B - Yes. It is. Indeed, the latest all applications possible they are all compatible with GenePix.

A - Now, can you just brief some of the applications of microarrays which people usually perform using these types of scanners and software.

B - So classically microarrays are being used for differential gene expression which usually people categorize in the gene expression format so as later develop has happened people went for SNP arrays as well CGH which is comparative genome hybridization to look at chromosomes, apart from that protein arrays, chip-on-chip arrays to look for the control gene they are also being focused and protein arrays are really catching up in many sense because the protein is actually the biological function that is what many scientists believe and I am also a strong believer. Now people are moving from the DNA to the real functional part how they directly affected. The major bottleneck was looking at the antigen-antibody reactions, the proteins are all being seen there is no general formula but now the field is catching up. So even at the level of protein arrays this is compatible. The incoming tissues arrays which are in infancy also now being compatible with GenePix Pro.

A - So whenever we talk about microarray people get a feel about it probably oligos, DNAs based microarray thing which we just talked about but slowly from this overall process from DNA to protein microarray now field is shifting in that direction but only

major limitation here is that how to generate protein content in the similar high throughput format people used to do for DNA arrays. So major success for DNA arrays was easily availability of DNA and stability of DNA and that's why the high-density arrays are possible from the DNA technology. But applying the same to the proteins becomes very challenging because purifying that many proteins keeping them functional and stable it becomes very challenging and that's why like in my lecture in class I have discussed some of the latest methods which people use for making protein arrays and we have talked about different type of technologies including how one can use the cell-free expression based system for generating the protein content which could be used for high throughput microarrays for the protein. So it's good to hear that there are different types of applications which one can use here starting from DNA to proteins. So now these scanners, what type of pixel resolution they can scan and what is the suitability for the scanning?

B - So our scanners range is from 5 micron resolution to 2.5 micron resolution. The basic thumb rule for the resolution is that the amount of size of your spot be protein or RNA that should be 10 times that of the resolution what you are scanning. So interestingly most of the DNA is less than 50 microns and most of the proteins usually stand at 200 microns so essentially the rule goes the 10:1. If you say 10 is a resolution and 1 is the spot size. So that is the major rule. So just to give you a brief out, say for an example if your spot of protein is 100 microns in size so what best, if you do a 10 microns basically you are getting 78 pixels so this is actually 1/10 if you do by 5 micron resolution all of the pixel will increase to 314 but the size of the image also increases so apart from making more than 10 times the size of the spot it does not really help at the level of resolution if you increase. The only thing you are increasing is size but the data quality just increase by few very minor changes so that is why the best ratio is 10:1 which 5-2.5 micron is normally suitable where people are making the slides.

A - Different scanners which aim for wide variety of scanning pixilation and resolution but I think some thumb rule which you mentioned is very widely applicable for any scanning.

B - Yes.

A - So the microarray data could vary in wide range so how to accommodate that variation while performing the scanning?

B - Yes. Essentially because they are all not present so the data could be right from 0 intensity which will be equivalent to the background to the very highest intensity. The best way to control the variations is looking at a photomultiplier tube, actually that is nothing once you do fluorescence so intensity is coming out of that so the very low expressing gene or the proteins will give you very low fluorescence intensity whereas high will give you very high.

A - Yeah. So just I will interrupt you here when people scan these microarray slides it always asks for PMT gain. Right? So I think that you are talking about photomultiplier tube and PMT gain. So just may be you can you clarify here how one can adjust that parameter of PMT gain?

B - So basic thumb rule there is scanning you don't want to see white spots. The white spot essentially means the saturation and to avoid saturation, exceptional cases for few controls people avoid that happening so the best way to look at photomultiplier tube condition so it ranges actually from something like 300-700 which helps one control how much photon multiplication can be done so that lowest can also be seen and highest does not become saturated. So this is how photomultiplier tube is very essential to control the different kind of variability within the chip, which happens during the biology.

A - So how to scan the best possible result? You would obviously like to change various parameters to obtain the best result. Right? So what are things which one should look for?

B - Yes. So the major variations actually come going to the different sources, one could be technical one could be biological and another one is going to the assay or chemistry. So the technical ones are in the form of hardware can be controlled at the level of PMTs and when you looking at signal to noise ratio, small play of PMT and very less with the

laser power and then with our own design of inverted scanning as well as the non-confocal chemistry we are able to achieve the best. So in short if by visibility you may not be able to look like what is happening to the images so in form, there here in the image which is being shown the which looks brighter to the eye may not be a true sense where you look at even the dull their background is very low without looking at the number if I see them I say this is dull that is bright but that is not true always. Because when you calculate reality based on the basis of what you have controlled on the level of PMTs and others. You quickly come to a conclusion that signal to noise ratio can be seen by eyes at level of PMTs which the laser balance can be seen so in this fashion while you are scanning the slide you can always look at that.

A - So one need to adjust the signal to noise ratio to achieve the very good image quality and obviously simultaneously need to perform PMT adjustment as well as other parameters.

B - True. As you go for the signal corrections, basically what you are trying to see spots not coming saturation and then you expect the image is going to be with least background so the information which is coming in the form of intensity become a true signal.

A - So how to quantify signal to noise ratio?

B - It is very easy to calculate once you have a signal and then when you have background which is already calculated so basically the intensity which is coming from expected spot and where there is no spot that particular intensity which is coming is naïve and there is another way of calculating the local background which is just surrounding by area of the spot that is called as a background, how we calculate-

Signal-Background: Background variation (= Standard deviation of the background).

A - Now one parameter which we see for the background correction basically looking for the scan line averaging so can you just elaborate on that?

B - Sure. Basically if you scan once so image is being collected so if you want to have multiple scanning done so that, that all intensities can be average upon and get to a signal value it is being seen that when you increase the number of scan usually people prefer 2-3, then average them so this helps this to reducing the signal to noise ratio in a fashion that you get a real good intensity coming up.

A - So can you just brief some of the very common features of this scanner because then only we can learn more when we talk about software interface?

B - Sure. So basic concept in the form of hardware design, the basic one is non-confocal optics and another one is focal-depth. The meaning of focal-depth is where our laser can focus in the form where it can acquire data and the depth is 64 microns which allows a wide variety of applications. If you want to use some slide based or a slide that is having a cover slip so the depth is little lower so you want to focus at different range so in this fashion non-confocal optics with focal depth of 64 is a major feature. Apart from that upside down that is inverted chemistry is actually what we are going to use in the form of printed slides being scanned directly and controlled for the variability I have discussed earlier and the resolution at 2.5 or 5 microns based on the system which is available can be also used. And real time PMT adjustment, PMT we have already elaborated the photomultiplier tube and the real time is at the level of scanning we can look at the data when the scanning live so in this we can control the PMT live in between the scan and then auto-PMT adjustment for those who are not real. How to control then we have given auto-PMT option. This auto-PMT option allows actually to get best output but GenePix Pro allows scan by itself so the best PMT-suitable for one application can then be seen immediately.

A - I am user of 4000 B so I would like to know little bit more some of the features of this 4000 B which is a classical scanner.

B - So GenePix classic 4000 B is classic because first introducing the 2000 but after that everybody started loving this machine owing to the fact it is having 6.5 minutes of simultaneous scan. The meaning of simultaneous scan is both lasers can act at the

same time. So you have very less time for scanning. Apart from that it has got two lasers, actually 635 and 532 nm which is classically used for cy3 and cy5 and they are compatible dyes. So in view of this we have standard green and standard red filters to accommodate all cy3 and cy5 applications. The laser power can also be adjusted in the form of 100 % to 33% to 10% so based on what application and what intensity you want based on that laser power can be adjusted. So adjusted focal position is from 50 to 200 microns so this allows you to focus in different ways so we can have a different slides compatibility coming up with the scanners- 5 micron resolution maximum it allows to go for and it can go upto 100 micron resolution. So non-confocal again it is using the non-confocal optical design. Any standard slide can be used for scanning so wide application is possible in that.

A - So data acquisition is always very important aspect because once we acquire a good image and good data then only one can analyze and obtain any meaningful information so can you elaborate how to acquire the image.

B - As we discussed in brief like the GenePix Pro software we are going to use for data acquisition. So basically there are different steps involved what in the usual process of bioinformatics. The first one comes is imaging so imaging is being done in the form of multi-imaging or single image tif format so we use tif format to create the images from and export these colored images in 24 JPEG format. So these JPEG format only allow one to see, the basic data processing will be done on tif format. So once GenePix Pro is allowed in a different sense of hardware controls then image acquisition can be done using number of lasers availability. It ranges from 2-4 and many times many applications use 1 laser as well. So based on the channel type, laser type you select how many to select for again on your chemistry basis. And this third party image alignment automatically being done so what this means is each particular array has got different blocks which in turn contains the features. So these features are actual genes pr proteins or the representative of biological material what you are checking. Now these has to be aligned with annotation information so this can be done by third party and the GenePix pro is doing by itself in the form of automation to do the job so

automated block and feature alignment is possible even the different sizes and shapes of the spot can be handled, right from the circular, square to irregular feature which can be handled at the level of image alignments.

A - So I think you have mentioned the good points about data acquisition because one need to look at various type of features and different type of analysis later on which one need to perform so data acquisition has to be very flexible in terms of some times the spots could be irregular, there could be different type of background, there could be different type of chemistries being used so I think by looking at some of these parameters one can acquire good image and then data acquisition can be further used for analysis. So maybe we can talk little bit on the analysis aspect?

B - So in any microarray after the image first thing comes alignment which we have just covered in the form of data acquisition because that is very important part to cover and apart from that you have a background subtraction so that you get a true signal coming in so GenePix Pro does help in background subtractions in the different format and also in the normalization features so you can see background subtraction can be done in the form of local, global, negative and morphological controls so this negative and morphological controls are subjected to the design of the slide type. Say negative controls in your slide there should be some spots which should not be bind to anything and leave blank whereas the local and global can be calculated in the general space where there are no spots available and the area which is not being spotted nearby your particular feature so in this fashion background subtraction can happen and then feature viewer and feature pixel plot so basically the major thing comes after the acquisition is visualization of the data and these visualization comes in the form of pixels and plots the graph so the graph helps us in understanding globally what is happening in short so it gives you a real image how the things are happening so this can also be done. And there are multiple ways of calculating in the form of ratio calculations after normalization or during the normalization of the data so analyses immediately after those involve few of the normalization process which GenePix Pro very well handles and other important feature is the flagging of the spot. In biology we see some spots are not really good or

because of some artifacts they should not be taken for the analysis so essentially we need to flag that spot as good, bad or absent so these can also be done by GenePix Pro software. Lastly normalization of the images and the different formats is also allowed to happen in GenePix Pro.

A - I think you rightly mentioned the need and importance of the control features because many times the quality of the experiment depends on how well your controls are performing? That applied to both positive and negative response and when you talk about DNA microarray technology obviously we are talking about very high density arrays here where lot of controls and lot of spots are built in place, many times when I talk about protein microarrays specially when we talk about functional arrays so we actually put different type of controls just based on that one particular experiment, for example I am looking for some biomarkers response I need to have certain positive controls some biomarker which need to light up on the array which will guide me as a positive control then if I am looking at immune response I need to have some sort of IgG and some of those type of control features which will guide me just how non-specific response could be so again empty spot we need some type of spots where there is no DNA or no protein is printed or no biological material is there so usually these type of controls are part of the design and that helps further for background subtraction as well as how good quality data we are obtaining.True. That is the major calculation for background work.

A - So let's talk about visualization like how one can really generate good image and which can be further used for analysis?

B - So once the image is being acquired so immediately scientist wants how my data has performed? As you have given very beautiful talk on how the controls works and specially few people use even bacterial controls which is not related to biology as such, totally they just want to see the assay controls so first they say my assay has worked or not so best way is to do for them to looking at the graphs which are being given differentiation and the level of background and can be also done at the level of the

histograms and the scatter plot is possible. These scatter plots can be plotted once against the channel types- laser 1 versus laser-2 or wavelength 1 versus wavelength 2, in classical say cy3 versus cy5. How these two things behave for me so again kind of graphs and also the images can also be exported to PDF as well as being visualized in GenePix Pro for your further screening for your QC applications.

A - So, I think It will be useful if we can actually see how the software works for performing the scanning but before we move on to live demo of the scanning performance may be we can briefly discuss the software in general and how the user interface looks like and if you can just elaborate on some of the features while one need to look on the scanning process I think that will be useful for us to understand when we actually look for the demonstration of the software.

B - True. The basic GUI which we are going to see in few minutes, actually contains three different areas the first one is image pleasure control so there you wanted to see which kind of laser I am going to use and second one is a different feature which are being used for controlling of the image and towards the right hand side we have a pen which is allowed or helpful in hardware control. So the basic one in the hardware control is first as we said we can look to the auto-PMT and other adjustments which is being done for that we use preview scan. In preview scan there are different tabs which allows you to do true scan or preview scan and based on the particular laser which scan you are using. If you start with a preview scan you decide which pixilation suits me whch different power of laser suits me once you are able to do these decisions made you can go for your own data scan. People prefer that because they are sometimes bleaching effect on the fluorophores so they want to avoid exposure for longer time so multiple scanning is avoided. So once you see a single wavelength scans if your application requires only one wavelength to be scanned, there you can select in the tab only a single wavelength scan and then show also a preview scan with data scan followed by. You can also do automation that once you do a preview scan and see it is all good then immediately go for that. But very rarely people use this because you see once and then you go ahead.

So it's like general scanning even if you want to scan a sheet first of all you want to preview it that how the overall image look like and then since you know your experiment, you know your requirements like what wavelength need to be used, what type of fluorophore you have used then one need to optimize and correct those things using the data scan which actual scan performed and then one need to review whole things and how the slide look like.

A - We can talk little bit about the software which even I am using GenePix Pro, how one can acquire picture if one has a system like 4000 B.

B - So there is a button towards the side where you have the control for the scanning time so as it is duo-channel that means two lasers are present in that. It is allowed to select whether you want to use one or two and then based on the one or the user application you select both the laser and then at the level of live scan you control for the PMT and whichever resolution you want to use for. So these are in the same software towards the right hand you see the pen where you have hardware control button there also you can look for different images which is now suiting for your own biological application.

I think it will be clearer when we will look on your software scan.

Sure.

A - OK. So when we are acquiring the image what the intensity histogram tell us while scanning is in process and even after the scanning is done? How one can really ensure that the scan is good and what type of balance one need to make in that?

B - So basically the preview scan when you are scanning your live data you can just switch on to histogram graph. There what it gives is how much green and red channels are contributing towards the intensity. So you really want that they are overlapping so they are really balancing, there could be a small variability in the beginning owing to the fact that they are just background and then the spots coming in. there you want that they are really overlapping after little bit of lag that is few seconds of lag that is it. So

once you are able to do then see and select whether yes this PMT is good for me so this Is the way you check which PMT is more suitable. So you select the PMT, look at them, see the overlapping where there is best overlapping without the saturation you want to go over those statistics.

A - I think this is the time we should really now move on to your software interface and see how scanning is performed by using this software.

B - OK. Let's switch on to the GUI.

A - OK. Pankaj so let's talk about software and how actually we can use this software for scanning from the hardware so that we can see it live that how one can control different buttons for acquisition of the good images.

B - Let's move on to the GUI.

B - So what you see now is GUI (graphical user interface) of GenePix Pro software, on top of it is in form of different tab buttons which allows you in different work groups for example 'image' allows you in different ways encountering of the image acquisition and histograms looks at how that image has performed so this is what we were speaking about in the earlier slide where you can see live kind of demo which is happening and then lab book actually gives you what all have you done in different stepwise so every movement of yours is in this software is being logged in and then analysis can be done in the form of batch form which allows multiple slides so that you do alignment and do analysis which can be performed as batch analysis. Once analysis is over results can be seen and the scatter plot can be plotted at the level of this graphical user interface. Once you are through you can look at the reports as well. So let's look at major function of imaging and how one can control for image acquisition.

A - I think that is important. Let's look at some of the buttons and how one can control those to get good image.

B - Let's quickly go through different kind of buttons here. Now the imaging can be at different wavelengths and preview can be done at 635 and 532 nm in a single laser based the wavelength can be done at 635 or wavelength at 532, even the ratio of the imaging how both has performed together can be looked at looking at button of ratio of imaging so this one allows you to see how the image is being done after the scan. You can look at one channel, preview channel or different channels. Now let's look at different tools which are available to you while or after the scanning so the major ones are here where you can move across the chip in form of this hand tool, plus indicates the zoom tool, and the other tools are- this is for the unzoom and you can also look at the whole image button. So once you have the image these tools becomes activated. So these two are for blocks and looking and controlling blocks and these ones are the features. So many a times what happens usually get a GAL files is a feature information file. Gal stands for gene array list so actually it gives the X and Y coordinates where each array usually is being present in the form of block which each block are in terms of the feature so these blocks and feature positions are being recorded in GAL file then the information or annotation is given to each spot so GAL file essentially contains the X and Y and the number of the columns and so the information of each spot how they are being annotated and placed on the chip so by chance if you don't know or you prepared by yourself these buttons here allows to make own block and create your own GAL file with the help of tool called gene array generator. So now let's look at the control button which is at the right side so first one is preview scan and then you also have a data scan. One stands for one wavelength so it allows you to take image from only one wavelength. And then you also have a multiple scan so you do a preview scan and then you do a multiple scan with this but you have other buttons which will light up when you have image in your hand and this is for the analysis. If you click this button the analysis will be performed after the alignment, this is actually open button so normally you have your file where you want to open and save your images. And this one is actually flag button as we discussed the features can be flagged, when image is available to you can look at good, bad or absent and you can give them different ratings. Here again looking at different zoom buttons which allows you to that which view you want to focus feature

name and feature IDs for which particular one you want to go for. This one here allows different work flow controls.

B - Now quickly go through a particular scan which is a simultaneous scan so both lasers would acquired at the same time so if I press on data scan button the image after putting your inverted image in the data hardware. It's scanning so you just see on the top which is very less visible as you see on the top. So let's try to zoom inside. If I put this button and allow zooming, you can see particularly how the scanning is happening. So you are looking at different image type so if I click on only one wavelength because it's live after scanning you can see it's going for ratio image scanning. So now quickly add a histogram you see it's started coming up because the scanning is going on live so it's stat reducing. Basically as we discussed it should be overlapping, my settings are looking very nice in this particular one.

A - So as scanning is progressing one need to keep looking at the histogram, to determine how cy3 and cy5 are aligned.

B - So how well aligned with the help of auto-PMT so cy3 and c4 you can adjust auto-PMT you can adjust laser power so that you can see this one.

A - So if we see some variation then we need to come back here and adjust these parameters so that they are super-aligned.

B - So in this fashion image acquisition is happening and we give you a power that in between the slides usually people keep barcodes and our system is compatible with internal barcode which is being done so that you can have multiple scans also being done so nowadays each slide is coming with multiple arrays because of the variable densities people are focusing on the custom type so this can also be done with the new software operation developments. So now as the scanning is being performed that's look like I have saved the image once I have saved the image I would liked to see how the different processes are being tagged. So say for example I have saved this image in the form of example this est so I just want to open an image which I have just saved so

basically as we discussed. Each particular array can be divided into the blocks so particular array of est contains four blocks and each block is having features so number of feature information is given in GAL so basic terminology is array, block and features so I need to align that GAL information of positions on top of this. So I have to put a GAL file and do my further analysis so what I am going to do now is open a GAL file which allows me for alignment so best feature of GenePix pro is its capability of finding features by itself, it's a little clumsy so just show how the zoom button looks like.

A - One needs to fine tune that alignment for overall proper image extraction.

B - Correct. The only thing you have to do here is just take block and allow it to move it to the first alignment and then what you can do is click the button over here which is for the align, align can be done in different ways I recommend to you the first which finds all feature, all blocks and do automated fashion so you click once you see software automatically finds all the features wherever the features are by chance absent there is some physical deformity it says it is not present or flagged as bad.

A - So once thing which is good here it is automatically adjust according to the spot size, the overall width it is adjusting because some spots are not uniform. Right? So it's making that correction here.

B - Correct. So as you can see you can move it but it does not affect life as long as you have kept burns and data is being stored but usually people ask me is it good idea if by chance move it once again so it's not a bad idea it takes a few idea to it. Once you have done this alignment let's look at these two slide which I said you can zoom out so you can see whole particular slide is now being scanned and aligned as well so it is very quick process which software performs very easily for doing the job. And once you have done this you can always hit a button of results. Now if I go to results it is empty if I click on results, results are being calculated and there 40 different columns which will be output in the form which GenePix understand different ways so just quickly looking at the major ones.   The major ones are here looking for this F measure intensity for different channel 635 or 532 and this background calculation is being done accordingly

in the same laser range so once you do corrections what happens you want to correct your intensity mean values with the values of the background so this is usually most important which people use for further calculations apart from ration of means or ratio of medians which can be calculated again and being presented to you in a different column format. So each column signifies different ones for example SD- standard deviation, CV- coefficient of variations and then different channels coming up. So in this fashion the results will be outputted if your image acquisition first controlling the part then allowing you to align and then do the analysis. So this is basic steps which anybody or everybody wants to do in microarray steps. So once you see the images the people end up in the form of results you have different columns available to you.

A - Yes. After that what is the next thing to look for like how good scattered plots are?

B - So there are different ways people want to visualize how my column because number make very less sense so the best way to look is scatter plot. Scatter plot allows you in many ways, what you plotting on X and Y axis and if here you see here I am just plotting actually towards F635 median over the F635 median so you are comparing two different channels how they have behave so essential rule is this mostly the microarray assumes all the chips are having the spots and which are not vary too much so you expect most of them to stand nearby the origin of the centre so this what you want to look at.

A - So I guess we talked so for about how to use hardware to scan a slide and by using software interface which GenePix Pro here to acquire all the data. Now the next challenge is to rally obtain some useful information from this whole data which we have already acquired.

B - True. As we discussed GenePix Pro is acquisition software and the molecular devices recommend acuity software for further data analysis which can be at the level of secondary or tertiary based on that. So you do statistics as well as visualizations on single as well multiple data to handle.

A - So Pankaj it was good to talk to you about how one can use DNA microarray or protein microarray to different type of slides and use the scanner to scan these slide by using this software which has various features depending upon experiment one can actually take a decision that what type of features to use for that scanning. I think it will be very good if you can just share the video which describe the overview of the whole process starting from the sample preparation followed by which image acquisition and image analysis. So if you can show us that video it will be great.

B - Thanks to molecular devices, they have provided me with this video which actually takes you from the basic process of biology in very brief to importance software parts and so also the hardware design which is being emphasized to the level of results so let's watch that video.

Video - Molecular devices introduces the worlds simplest, most reliable automatic microarray slide scanner. Now you can walk away from scanning while GenePix autoloader 4200 AL automatically loads, scans, analyzes and saves results for up to 36 slides. The autoloader accommodates microarrays on standard glass microscope slides labeled with upto four fluorescent dyes. These microarrays can contain just a few hundred spots or tens of thousands of spots representing an entire genome. As many as 36 slides can be loaded in convenient slide carrier as the carrier is inserted into the scanner sensors detect the location of each slide indicated by a blue bar on the slide carrier map. On the batch scan tab in GenePix Pro you have complete flexibility to define the most appropriate and analysis parameters for each slide or for groups of slides. You can also choose to automate scanning, analysis and file saving steps. Enter an email address and GenePix pro will notify you remotely when your batch is complete.

Video - Using the defined scanning parameters the precision robot arm leads in the action and moves to the first slide. Our unique never let go wipers securely clamp the slide and carry it to the scanning area. A barcode reader records the barcode. And then the slide is positioned for scanning. The GenePix autoloader 4200 AL can be configured with up to 4 lasers. A neutral density filter wheel can be used to attenuate the laser

power if necessary for especially bright samples. The laser excitation beam is delivered to the surface of microarray slides, the beam scans rapidly across the short access of the slide as the robot arm moves the slides more slowly down the long access. Fluorescence signal emitted from the sample is collected by a photomultiplier tube. As the scan proceeds, sensors detect any non-uniformity in the slide surface and the robotic arm adjust the slide position accordingly to ensure the array surface is always in perfect focus. Each channel is scanned sequentially and the developing images are displayed on the monitor. The multi-channel tif images are saved automatically according to file naming conventions specified by the user. After the slide has been scanned, the precision robot arm replaces it safely in the slide carrier before picking the next slide. As each slide is scanned a list of each saved image with associated settings and analysis files accumulates in the batch analysis tab until the batch is complete. GenePix pro automatically finds the spots and calculates up to 108 different measurements for each spot, results are saved as GenePix results or GPR file, GPR files can be saved automatically to the acuity database for statistical analysis, clustering and other advanced investigations.

A - OK Pankaj. It was very useful discussion with you Pankaj for knowing more details about various type of features one need to look for to obtain very good images because that is the most important part for doing the microarrays. Once you have acquired good images then only one can do good data analysis from that so thank you for coming here and sharing your experience for overall scanning and data scanning. Thank you.

A - So in summary today we talked about microarray scanning and processing. We had a discussion and demonstration to over basic concepts to microarray image scanning and processing. One need to look at various parameter while scanning images and image processing because that is very crucial for doing the further data analysis before you want to obtain any biological meaningful information you need to very carefully process the image and then further perform the data analysis. In the next lecture we will continue our lecture on microarray workflow and how to analyze the microarray data obtained from these images which we have discussed in today's lecture. Thank you.