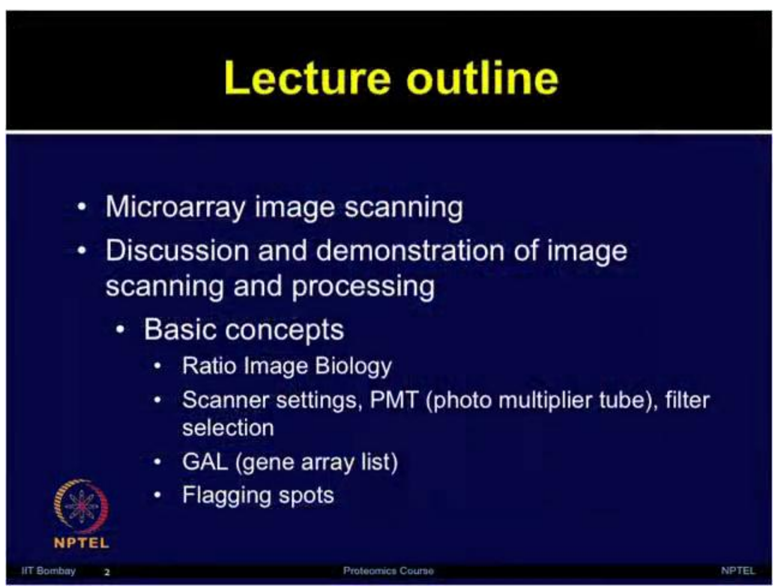
Proteomics: Principles and Techniques
Prof. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay

Lecture No. # 31
Microarray Work-Flow: Image Scanning and Processing

Welcome to the Proteomics course. In today's lecture, we will talk about Microarray Workflow, the Image Scanning and Processing; this is in continuation of our discussion on microarray workflow.

(Refer Slide Time: 00:38)

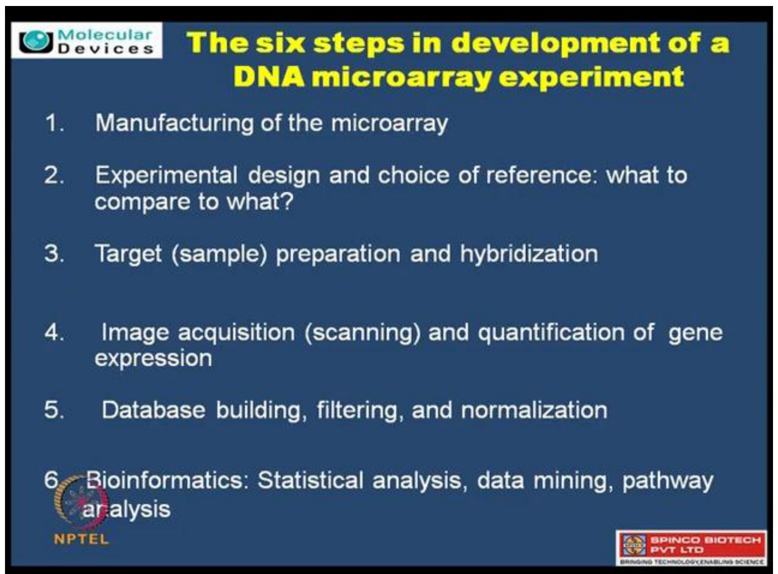


This is my pleasure to introduce Mister Pankaj Khanna, manager-application support from Spinco Biotech Private Limited. Today, we will be talking about GenePix a microarray scanner, which is from molecular devices, and Spinco is distributor for the same. So, we will have a discussion about microarray 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 the software interface so that you can get a better feeling about how this whole process is controlled and perform. So, Pankaj, welcome to discussion about microarray scanners and software.

Thank you, doctor Srivastava, it is a pleasure.

At Spinco Biotech, we are having molecular devices as distributors for last many years and we are dealing with the scanners for many years at the level of microarrays. As you have already described that this is used in the form of microarrays, we are dealing from the image acquisition to the level of data analysis for the microarray applications. So, let us go through little bit on what is GenePix actually can do in a form.

(Refer Slide Time: 02:05)



So, that 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 talking more on the protein microarrays and things, but this whole technology got initiated from the DNA microarrays. So, just, so the students have some idea about how the DNA microarrays are performed, just give some brief steps about overall DNA microarray technology and also like, how and what GenePix can help in the microarray process. Sure.

So, basically the DNA microarray steps are having divided into six basic steps, starting from the manufacturing of the microarray. So, based on the slide and biology, the manufacturing of the slide is done. See, if you want to go for the RNA expression chip, you have RNA expression chip made. Right.

So, if you are using a protein expression chip, you are using a protein expression chip. So, after the design is being decided, the next step is the experimental design which usually scientist think based on the biological questions what they are asking. Right.

So, based on that, they choose which reference, which controls to be studied and group that. Once you are ready with this two, the next thing comes is the sample collection and the target or the sample preparation, which then can be hybridized with the preplan chip, which now is ready to go for the scanning in the form of image scanners. Right.

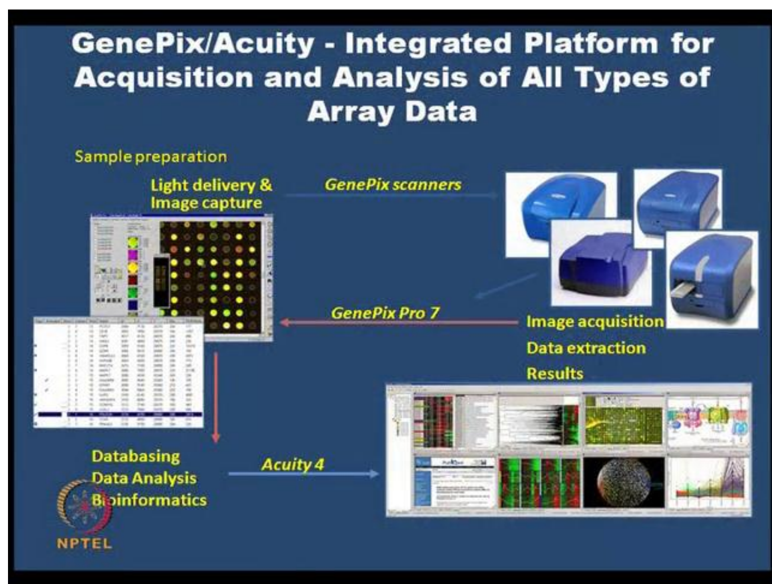
So, the fourth step involves, actually the hardware beginning up, where we as a vendor from the molecular devices come in place and this data once scanned in the form of raw data, now can be meet data building, filtering as well as the normalization followed by the last steps of bioinformatics and biostatistics which help in the data analysis. So, in this fashion, basic these six steps are important for any microarray experiment.

So, just to clarify, summarize here, I guess it is the content which is just different, whether it is a DNA or protein, but once we have the chips in hand, then entire assay and the whole procedure is very similar. So, manufacturing of the microarray will be different for each of the content type, then experimental design one has to choose depending upon the content and the biological question what they ask. And once these things are in place, then one has to chose, what type of biological samples or clinical sample they want to probe, and then followed by perform the hybridization steps; at then once the experiment is performed, then one need to acquire the images and that is where these hardware and scanners come into the play. Yes.

Followed by we need to do the data analysis for these and then the statistical part comes into the play where we need to identify the meaningful information from these experiments. So, these scanners are user friendly and the sort of, even the cost wise and overall performance wise, they are quite good, they are used worldwide. I had been using it in my research earlier in the different places and also I am continuous user of this scanner. So, I am quite familiar with the overall configuration and the hardware part for GenePix scanners.

So, as you are using 4000 B which is actually the, as you said, the world best the hardware in the form of Cy 3 and Cy 5 dyes been used, apart from that people use for different applications as well. Right.

(Refer Slide Time: 05:33)

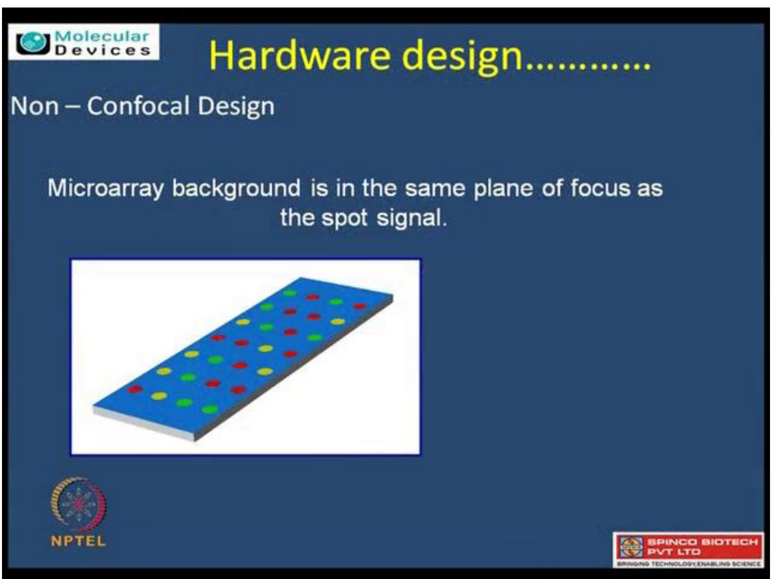


So, can we talk a little about the platforms which are used for acquisition and analysis of these microarray data. Sure. So, in the form of hardware as we have discussed, there are number of different hardware possibilities in the form of GenePix, right from 4000 B to 4300, 4400, these are actually doing 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 as GenePix is being attached with a software called GenePix pro software, which helps in understanding how a 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, the hardware is being attached with the data acquiring software called GenePix pro and this GenePix pro software is now give the results to acuity which can be further analyzed for statistical analysis. Yes, obviously, you need one software which can help in the performing, scanning process and then once the images are required, then you need to obtain some meaningful information from those images and that is why I think software like acuity can help what they are. Yes.

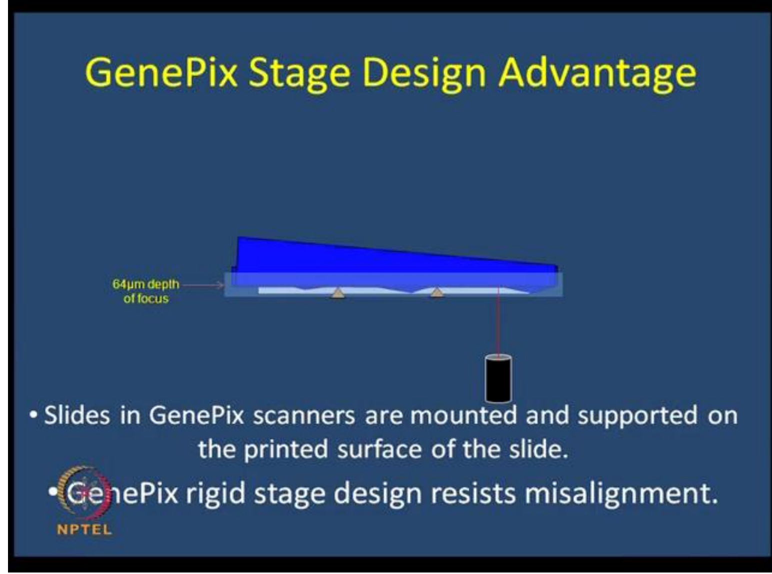
So, which technology is better for microarray and what do you think is more unique feature of using this hardware? So, GenePix microarray asset is based on two special designs, one of them is non-confocal design of hardware and second is inverted chemistry. So, what is this non-confocal design?

(Refer Slide Time: 07:24)



So, non-confocal design is like where you have a spots in the same plane where the scanning is done is called as a 2 D dimensional. Say, when you go for a tissue arrays or analysis where it is very very higher, so in the form of spots size, especially if the level of imaging of scanning images. So, there at the level of microscopy, people use a technology of confocal microscopy so that they focus a 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 are in the same plane, which is why it is non-confocal chemistry which just resides in the same plane of the slide. So, that is why we choose non-confocal chemistry attached with the inverted chemistry of scanning which helps in the best results in the form of signal to noise ratio. So, what are some of the advantages of this design of the GenePix stage?

(Refer Slide Time: 08:37)



Right. So, as we discussed non-confocal coupled with the inverted inverted scanning, how does it help is, usually we see in the glass slides, they are lot of small deformities which our eyes cannot see. So, what we do is, we invert the slide so that it goes to the level of the same scanning and they are supported at the level of the edges with the controllers which see the deformities at the glasses. Ok.

So, what it helps is, it helps in looking at the deformities and directly allow the scanning to happen on the phase of the scanning area. So, what happens is, because of the confocal design attached with this inverted chemistry, we get best signal to noise ratio so that your data is more valid for any analysis. So, this process can actually resist the, any misalignment which may happened due to the design of the slide. Very true. And overall scanning procedure. Very true, very true.

So, other different type of slides which are compatible for these type of experiments? So, as you have already said these systems are used worldwide and in different different academic institutes, people use at the various academic as well as commercial vendors. So, ranging from different applications. Right.

To different vendors, this support to name few say for example, Agilent, Nimblegen, Corning, these slides are all compatible with them which usually supplies in the form of either the RNA or the DNA chip, invitrogen which supplies the protein chip is also compatible with GenePix pro. So, apart from lot of (( )), all chips are also being

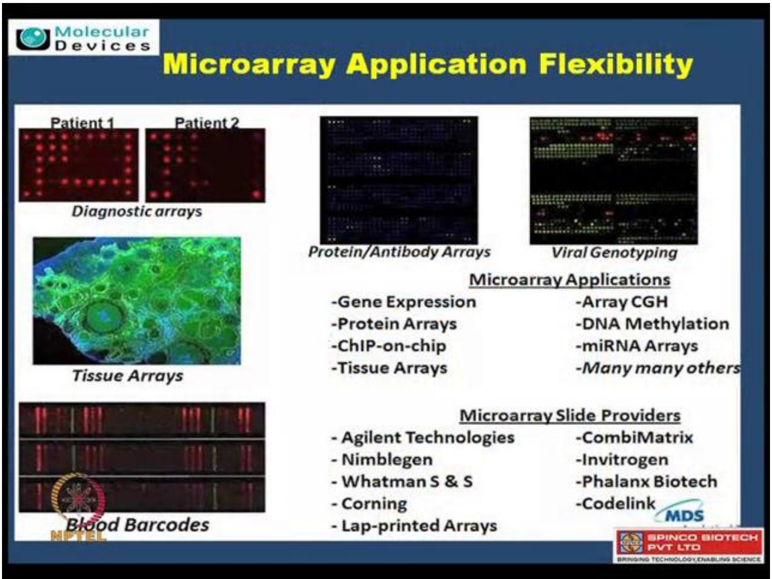
supported and incoming tissue arrays which are there, they are all supported with this particular technology as well.

So, it is a very useful to have a platform where one can use actually the slides from different commercial resources, otherwise for example, if one wants to use the same scanner for doing DNA array as well as protein arrays may it will become difficult. But as you mentioned like, we can also use the proto arrays from invitrogen as well as DNA array somewhere is type of commercial places. So, I think it just show that this platform is quite robust for wide applications.

(Refer Slide Time: 10:38)

Yes it is, indeed if the latest all of applications possible, they are all compatible with GenePix, because of the pixilation at 2.5 also which is now available. Now, can you just brief some of the applications of the microarrays which people usually perform by using these type of scanners and the software?

(Refer Slide Time: 11:13)



Yes, so classically microarrays are being used for the differential gene expression which usually people categorize in the gene expression from it. So, later as the development has happened, people went for SNP arrays as well as a CGH which is called as comparative genome hybridizations. They are used to look at the chromosomes, apart from that protein arrays, chip-on-chip (( )) control genes, they are also being focused and the protein arrays are now really catching up many sense, because the protein is actually the biological functions what many scientists believe.

(( ))

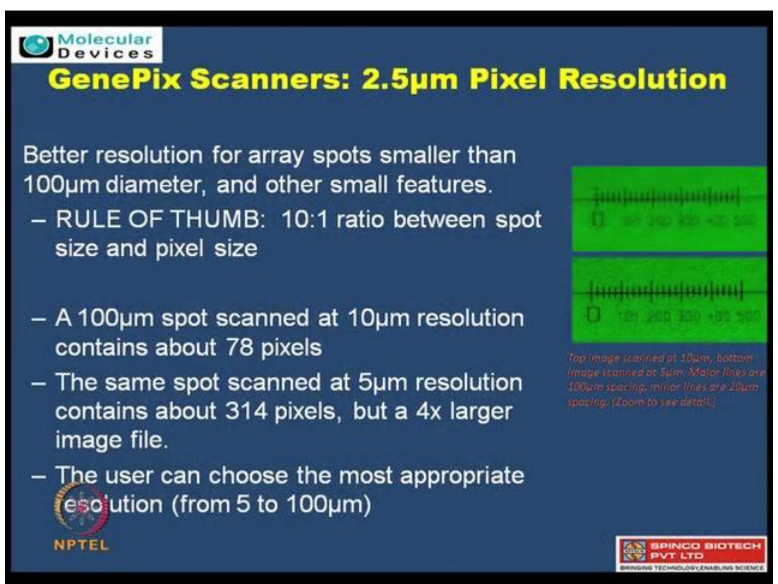
And I am also a strong believer, so now people are moving from the DNA to the real functional part, how they are directly affected that with the major bottleneck was looking at the antigen antibody reaction. Now, if the proteins are all being seen there is no general formula, but now the field is catching up. So, even at the level of protein arrays, this is compatible, the incoming tissue arrays which is really infancy is also now being compatible with GenePix pro.

So, although like whenever we talked microarray, people get to feel about it is like probably oligos or DNA based microarray thing which we are talking, but slowly now this overall process from DNA microarray to protein microarray, now field is shifting in that direction. But only major limitation here is that how to generate that content, protein content in the similar high-throughput format, what people used to do for the at DNA

arrays. So, major success for DNA arrays was, they easy availability of DNA and the stability of DNA and that is why the high density arrays are possible from the DNA technology. But applying the same at the protein becomes very challenging, because purifying that many protein, keeping them stable and functional, it becomes very challenging.

And that is why, like in my previous lecture in the class, I have discussed some of the latest methods which people can used 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 the high-throughput microarrays for the proteins. So, yeah, it is good to here that there are different type of applications which one can use here starting from DNA to proteins. Yeah. So, now these scanners, what type of pixel resolution they can scan and what is the suitability for scanning?

(Refer Slide Time: 13:26)



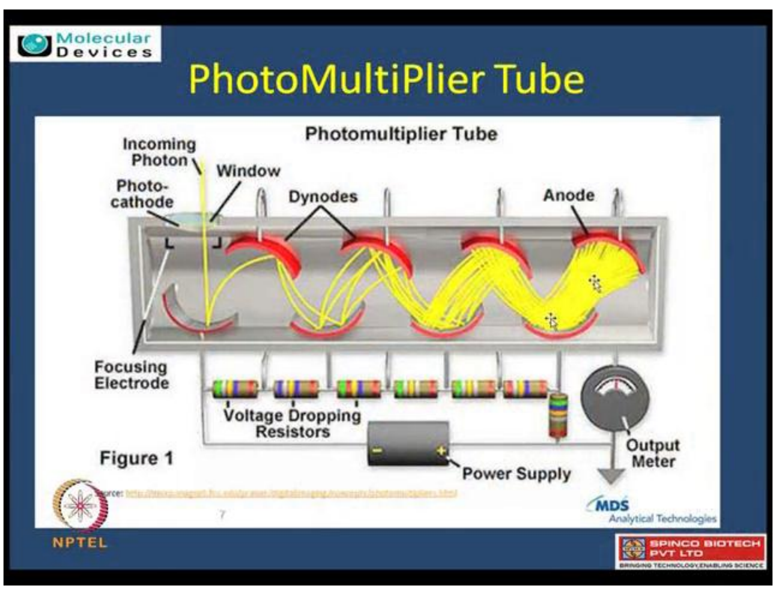
Right. So, our scanners ranges from 5 micron resolution to 2.5 micron resolution, the basic thumb rule for all the resolution is that the amount of your size of your spot (( )) protein or RNA Right. That should be 10 times that of the resolution what you are scanning that. So, in interestingly, most of the DNA is less than 50 microns and most of the protein usually stand at 200 microns. So, essentially the rule goes that 10 is to 1, if you say 10 is resolution and 1 is the spot size. So, that is the major rule, so just to give

you a brief out, say for 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. Hmm.

So, this is actually 1 by 10, if you do by 5 micron resolution although the pixilation when increase to 314, but the size of the image actually also increases. So, apart from making more than 10 times size of spot, it does not really help at the level of resolution if you increased, only thing is you are increasing the size. But the data quality is just increase by few, very very minor changes, so that is why the best rule is 10 is to 1 which 5 to 2.5 micron usually is more suitable. Right. Where the people are making the slides.

So, there are different scanners which aim for wide variety of scanning, pixilation and resolution, but I think what some of the thumb of rule which we were mentioned is very widely applicable for any scanning. Yes. So, the microarray data could vary in wide range. So, how to accommodate that variation while a performing the scanning?

(Refer Slide Time: 15:15)



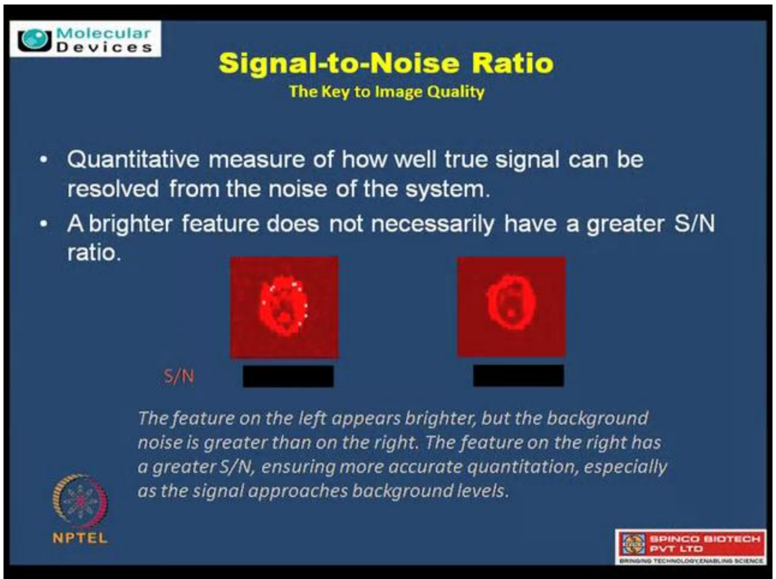
Yes, essentially because they all not present. So, the data could be right from the 0 intensity which will be equivalent to the background to the very highest intensity, to best way to control the variations is looking at a photomultiplier tube, actually that is nothing but, once you do a fluorescence,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

Yes, I just when interpret to here, when people scan these microarray slides, it always as for the PMT gain right. So, I think that is what you are talking here, photomultiplier tube and the PMT gain. Yes. So, just may be you can clarify here like, how one can adjust that parameter of PMT gain. So, the basic thumb rule there is, it any scanning you do not want to see a wide spot, the wide spot essentially means the saturation and to avoid saturation, exceptional cases for few controls, people avoid that happening. Ok.

So, the best way is to look at the photomultiplier tube condition. So, it ranges actually from something like 300 to 700 which helps one control, how much photon multiplication can be done. So, the lowest can also be seen and highest does not become saturated. So, this is how the photomultiplier tube is very very essential to control the different kind of variability within the chip which happens due to the biology.

So, how to scan a slide for best possible result? You know obviously, like to change various parameter to obtain the best result right. So, what are the things which one need to look for? Yes, so the major variations actually come into the different sources, one could be technical, one could be biological and another one is owing to the assay or the chemistry. Right.

(Refer Slide Time: 17:14)



So, the technical ones in the form of hardware can be controlled at the level of PMT's and when you are looking at signal to noise ratio, the small play of PMT and very less with the help of laser power and then with our own design of inverted scanning. Right.

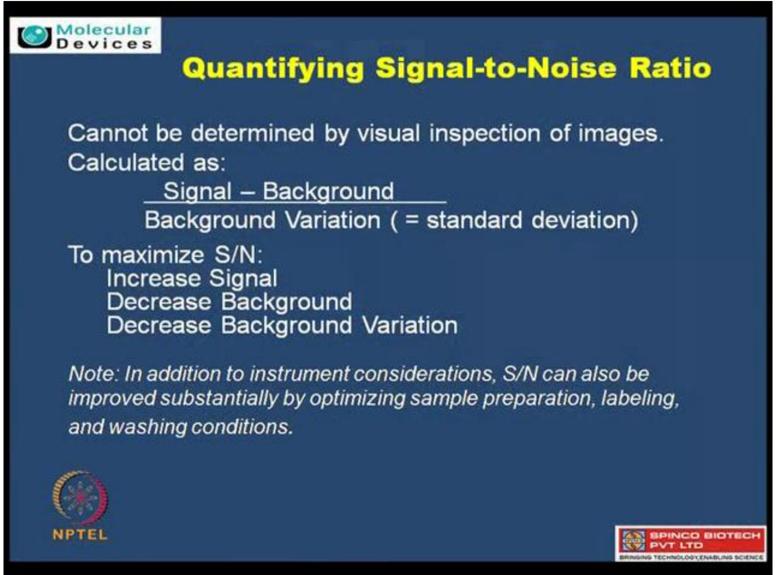
As well as the non-confocal chemistry, we are able to achieve the best. So, in short, if by visibility you may not be able to look like, what is happening to the images. So, in the form, there here in the image which is being shown the which which looks brighter to the eye, may not be a true sense. Right. So, when you look at even the dull, there the background is very low; without looking at numbers if I see them, I say this is dull that is bright, but that is not true always. Yeah.

Because when you calculate reality based on what you have controlled at the level of PMT's and others, you quickly come to a conclusion that signal to noise ratio can be seen by eyes at the level of PMT which the laser balance can be seen. So, in the fashion while you scanning a slide, we can always look at that.

So, one need to definitely adjusts the signal to noise ratio to achieve the very good image quality and obliviously, simultaneously need to perform the PMT adjustment as well as other parameters. True, so as as you go for the signal corrections, so basically what you are trying to see, same (( )) not looking saturations. Right. And then you expect the image is going to be with least background. So, the valid information which is coming in the form of intensities becomes a true signal.

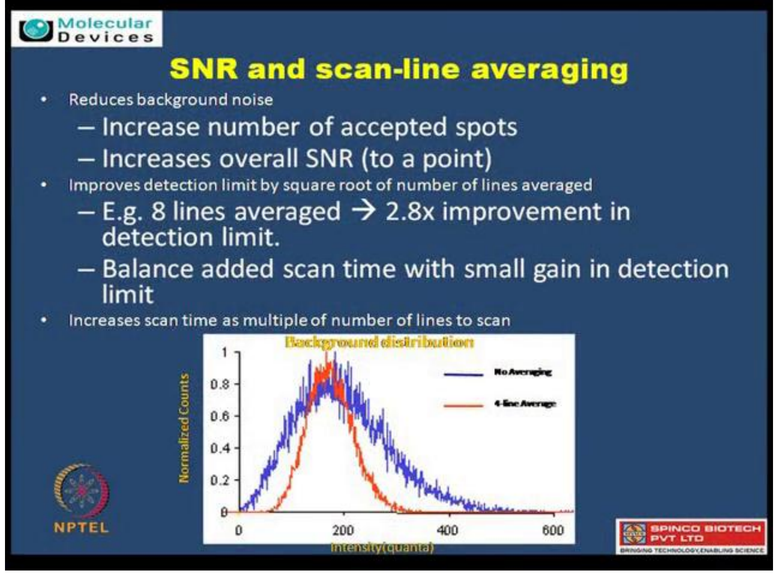
So, how to quantify the signal to noise ratio? It is very easy to calculate, once you have a signal and then when you have a background which is already calculated. So, basically signal is the intensity which is coming from the expected spot and where there is no spot that particular intensity which is coming is (( )).

(Refer Slide Time: 19:07)



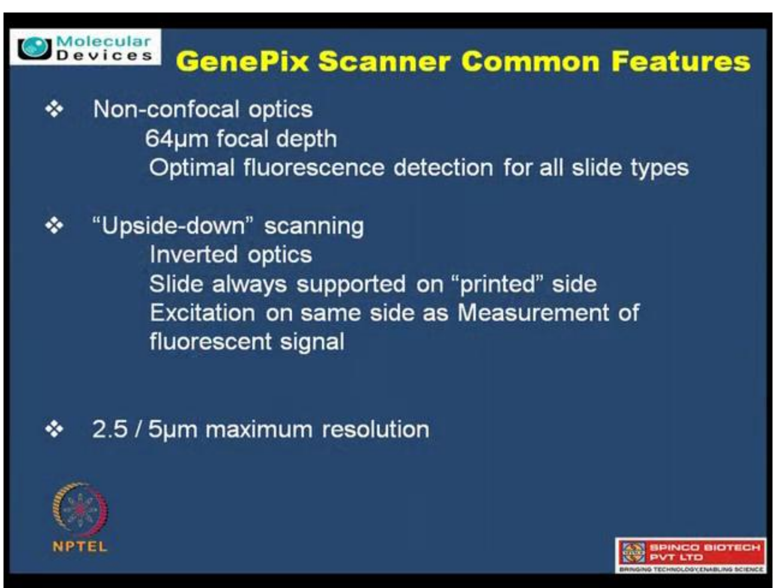
And there is another way of calculating a local background which is just nearby surrounding area of the spot. Ok. That is called as background, so we calculate signal minus background. So, we divide that with overall background deviation that is standard deviation of the background, that will help calculating a signal to noise ratio. Now, one parameter which we see for the background correction is basically looking for the scan-line averaging. So, can you say elaborate on that?

(Refer Slide Time: 19:40)



Sure, so basically if you scan once, so image is being collected. So, if you want to have multiple scannings being 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 preferred 2 to 3, so then average them. So, this helps in reducing the signal to noise ratio in a fashion that you get a real good intensity coming up.

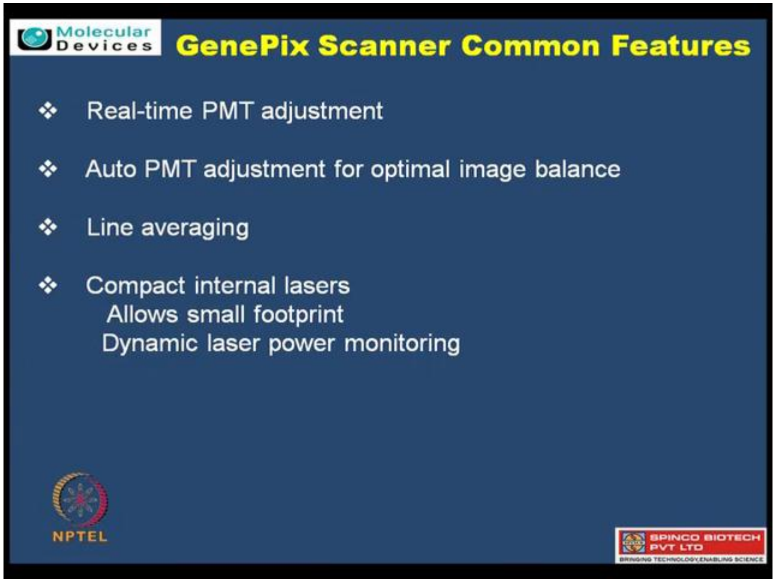
(Refer Slide Time: 20:06)



So, that is how you can brief some of the very common features of this scanner, because then only we can learn more when we talk about a software interface. Sure. So, the basic concepts in the form of hardware design, the basic one is non confocal optics and another one is the focal depth. The meaning of focal depth is where our laser can focus in the form where it can get our acquired the data and the depth is 64 microns which allows a wide variety of applications possible.

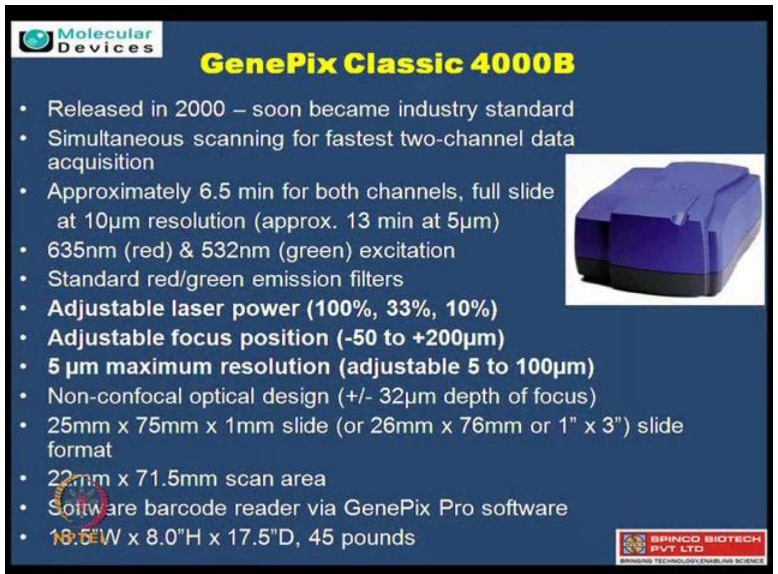
So, if you want to use some slide based slides that is having a cover slips so that the depth is little bit you know kind of lower, so you want to focus at the different range. So, in this fashion, the non confocal optics with a focal depth of 64 is our major feature. Apart from that up down, upside down that is inverted chemistry is actually what we are going to use in the form of printed sides being scanned directly and controlled for the variability as discussed earlier. And the resolution at 2.5 or 5 microns based on the system which is available can be also used.

(Refer Slide Time: 21:19)



And real time PMT adjustment, PMT we have already elaborated the photo multiplier tube. Right. And the real time is at the level of scanning, so we can look at the data when the scanning is live. So, in this form, we can control the PMT live in between the scan and then auto PMT adjustment for those who are not real burse of looking at how to control them, we have given auto PMT option. This auto PMT option allows (()) to get the best output, but GenePix pro allows the multiple scan by itself, so the best PMT suitable for one application can then be seen immediately. Now, I am user of 4000 B, so I would like to know little bit more about some of the features for this 4000 B which is more classical scanner.

(Refer Slide Time: 22:16)



So, GenePix classic 4000 B is called classic because it is like first introduced in 2000. But after that, everybody started loving this machine going to the fact that it is having 6.5 minutes of simultaneous scan, the meaning of simultaneous scan is both laser scan at a same time. So, you have very less time for scanning, apart from that it has got two lasers actually 635 and 532 nanometers which is classically used for Cy 3 and Cy 5 and their own compatible dyes. (( ))

So, in view of this, we have standard green and standard red emission filters to accommodate all Cy 3 and Cy 5 applications, the laser power can also be adjusted in the form of 100 percent to 33 percent to 10 percent. So, based on what application and what intensity you want best, the laser power can be adjusted as well, so adjusted focal position in from minus 50 to 200 microns, so this allows you to focus in a different ways. So, we can have a different slice compatibility coming up with the scanners, 5 macron resolution maximum it allows to go for and it can go up to 100 microns of resolution. So, non confocal again, it is using the non confocal optical design and any standard slide can be used for this scanning, so it gives wide applications possible in that.

So, data acquisition is always a very important aspect, because once we acquire a good image and good data, then only one can analyze to obtain any meaningful information. So, can you elaborate how to acquire the image? As we discussed in brief like the GenePix pro software, we are going to use for data acquisition. So, basically there are

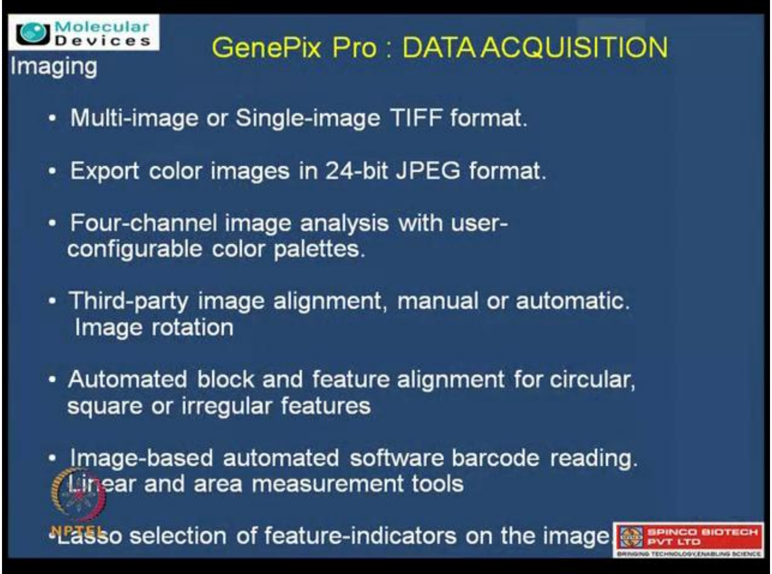
different steps involved, what in the process of bioinformatics, the first one comes is imaging. Right.

So, imaging is being done in the form of multi image or single image tiff format. So, we use tiff format to create the images form and export these colour images in 24 bit jpeg formats. So, these jpeg formats allow one to only see, but the basic data processing will be done on tiff format. So, so once GenePix pro is being allowed in a different cells of hardware controls, then image acquisition can be done using the number of lasers availability, it ranges from 2 to 4 and many times few 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 and manual automatically being done. So, what this mean is, each particular array has got different blocks which in turn contains the features. So, these features are actual genes or the proteins or the representative of the biological material what you are checking.

Now, these has to be aligned with the annotation information, so this can be done by a 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 circular, square to irregular features which can be handled at level of image alignments.

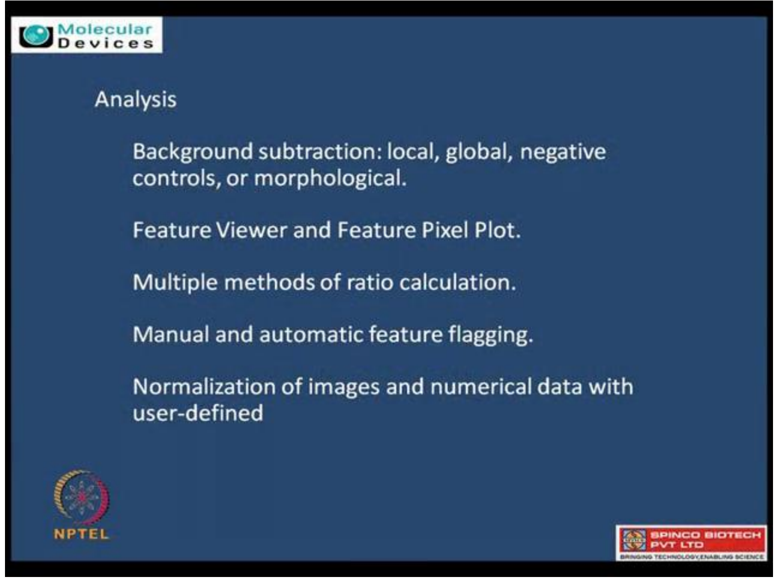
(Refer Slide Time: 25:46)



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

Sure, so in any micro array after the image, first thing comes is the alignment which we have just covered in the form of data acquisition, because that is a very important part to cover.

(Refer Slide Time: 26:42)



And apart from that, you have a background subtraction so that you get a true signal coming in. So, GenePix pro does help in the background subtractions in a different format and also in the normalization of the features. So, if you see the background subtraction can be done in the form of local, global and negative or morphological controls. So, these 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 bind to anything and leave blank.

And whereas, the local and global can be calculated in the general space where there are no spots available and the 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 plots.

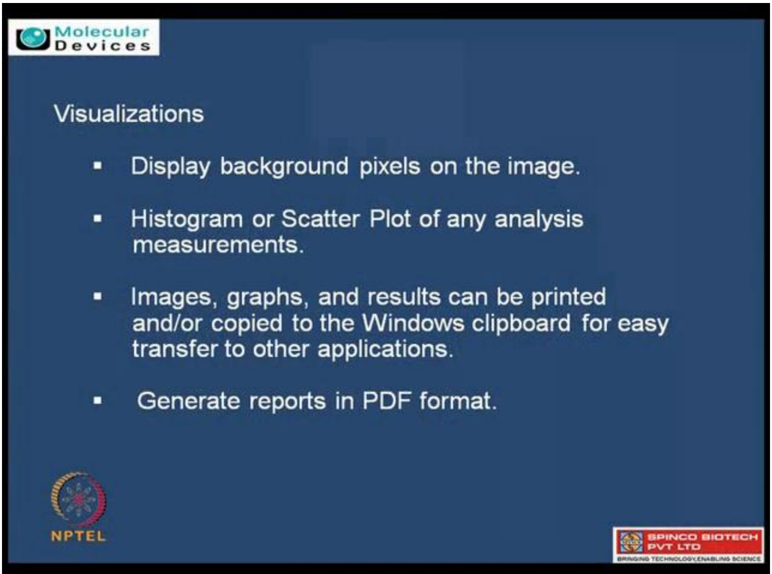
So, basically the major thing comes in the acquisition is after acquisition is, visualization of the data and these visualizations comes in the form of pixels and the plots the graphs. So, graphs help us in understanding globally what is happening in short, so it gives you a real image, how the things are happening. So, this can also be done and there are multiple ways of calculating in the form of ratio calculations after normalization or doing the normalization of the data, so analysis immediately after those involves few of the normalization process which GenePix pro can very well handle and other important feature is the flagging of the spot. In biology, we see some spots are really not good or because of some artifact, they are not supposed to be taken for analysis. Yeah.

So, essentially we need to make that spot as flag as good bad or absent. So, this can also be done with GenePix pro software, lastly normalization of the images and the different formats is also allowed to happen in GenePix pro. I think you rightly mentioned the need and the importance of the controlled features, because many times like a whole experiment is the quality of the experiment depends on, how well your controls are performed that applies to both negative and the positive control. And when you talk about DNA micro technology, obviously we are talking about very high density arrays here, where lot of spots and lot of controls are already in inbuilt in place right.

Many times when I talk about protein micro arrays and especially when we talk about functional type of arrays, so we actually put different type of controls just based on that from particular experiment. Further, if I am looking for some biomolecular response, I need to have certain positive controls some biomolecular 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 control features which will guide me it is just how non specific the response could be. So, again some empty spots, we need some type of spots where there is no DNA or no protein is printed there, so no biological material is there. So, lot of different type of control features are usually part of these array design and that helps further for background subtraction as well as looking at how good quality data we are obtaining. True, so that is a major calculations for the background (( )). Right.

So, let us talk about visualization like, how one can really generate some good image and which can be further used for analysis.

(Refer Slide Time: 30:13)



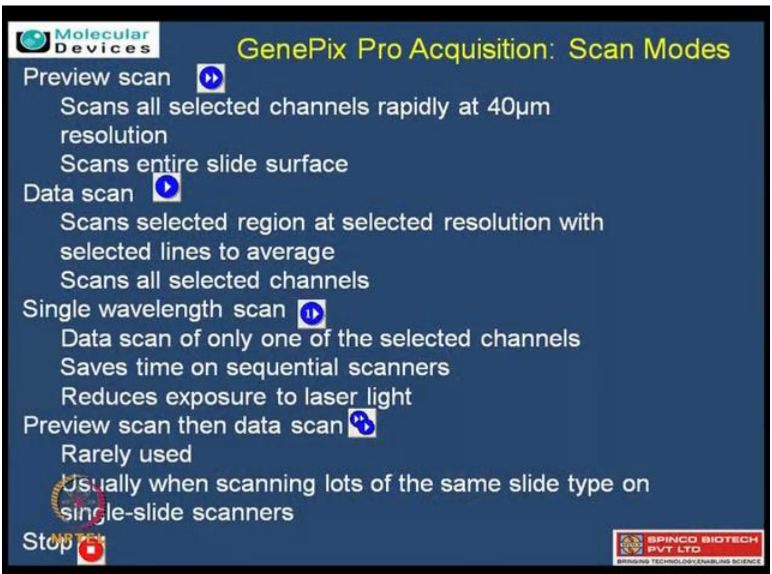
So, once the image is being acquired, so immediately scientist wants how my data has performed; as you have given a very beautiful talk on how the controls whirls and specially few people use even the bacterial controls which is non related to the biology as such, totally they want to see the just the assay controls. So, first they say whether my essay has worked or not, so best way to do for them is looking at the graphs which have been given differentiation in the level of backgrounds and though also can be done at the level of different histograms and the scatter plots possible. Right.

And these scatter plots can be plotted once against the channel types in laser1 versus laser 2 or wavelength 1 versus wavelength 2 in classical say Cy 3 versus Cy 5, so how these two things are behaved for me. Right. So, again different kind of graphs and so, also the images can also be exported to PDF as well as being visualized in GenePix pro for your further screening for different QC applications.

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

look at the demonstration for the software. True, so the basic GUI which we are going to see in few minutes actually contains three different areas, the first one is a image laser controls.

(Refer Slide Time: 31:56)

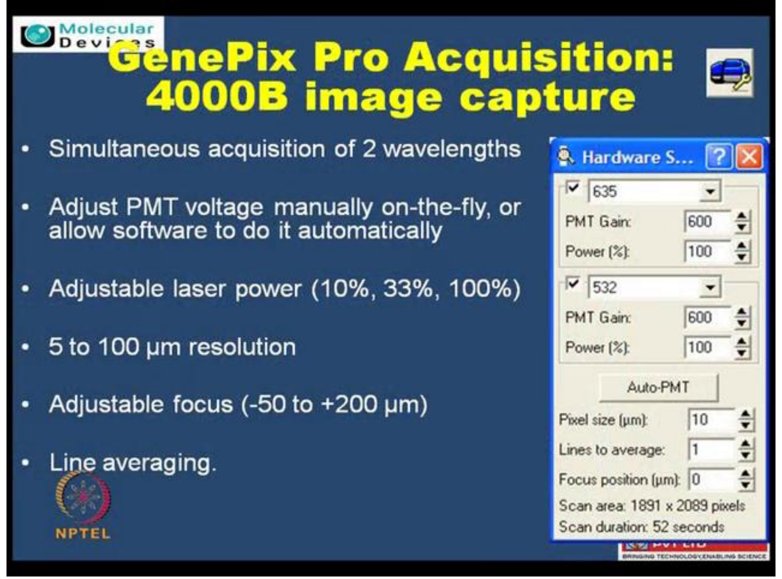


So, there you wanted to see which kind of laser I am going to use and second one is different features which are used for the controlling of the image. And towards the right hand side, we have a pain which is allowed or helpful in hardware controls. So, the basic one is in the hardware control is first, as we said that we can look at auto PMT and other adjustments which is being done, for that we use preview scan. So, preview scan and there are different tabs which allows you to go for the true scan and the preview scan and based on the laser which particular scan you are using. So, like for example, if you start with the preview scan, you decide on which best pixilation suits me. Right. Which different power of laser suits me.

Once you are able to do this decisions being made, you can go for your own data scan, people prefer that, because they are sometimes a bleaching effect on the flouro floor. So, they going to avoid the exposure for a longer time. True. So, multiple scanning is avoided. So, once you see even a single wavelength scan, so if your application requires only one wavelength to be scanned, there you can select in the tab only a single wavelength scan. Right.

And then so, also a preview scan with the data scan followed by, you can also an automation that once you do a preview scan you see that it is all good, then immediately it can go for that. But very people use that, because you see once and then only you want to go ahead. So, its like any general scanning, even if you want to just a scan a sheet right, first of all you would like to preview it that, how you you you just want to get a glimpse of the process like, how the overall image looks like. And then since you know your experiment, you know your requirements like, what wavelength need to be used, what type of fluoro floors you have used, then one need to optimize and correct those things within the data scan which is actual scan performed and then after that, one need to review that whole thing and then how the slide looks like. True. Maybe, we can talk little bit on the software which even I am using GenePix pro software and how one can acquire image if one has a system like 4000 B. Right.

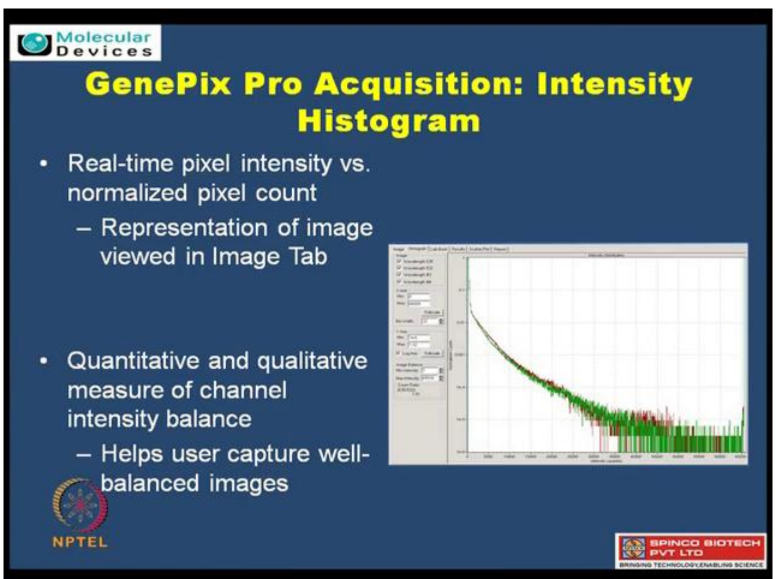
(Refer Slide Time: 34:01)



So, there is a button towards the side where you have a, having a control for the scanning time. So, as it is a dual channel that in the two lasers are present in that, it is allowed to sell right whether you want to use 1 and 2. And then based on 1 or the user application, you select both the lasers and then at the level of light scan, you control for the PMT and the (()) of resolution you want to use for. So, these are within the same software towards the right hand side, you see the (( )) where you have a hardware control button, there you can also look for all these different images which is being now suiting for your own biological applications. I think it will be more clear when we are actually looking at the,

your software interface. Sure. So, when we are acquiring the image, what the intensity histograms tell us while the 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? Yes, so basically in the preview scan, when you are scanning your live data, you can just switch on to the histogram graph. Right.

(Refer Slide Time: 35:19)

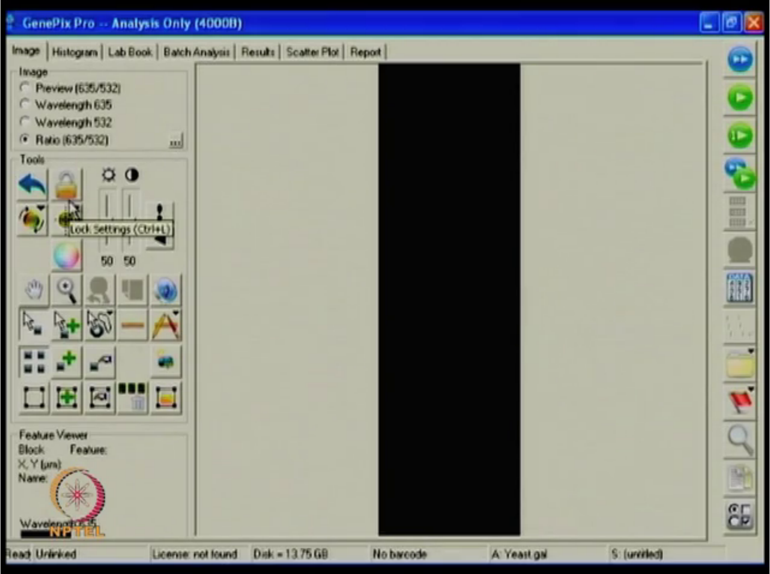


There what it gives is, how much red and green channel are contributing towards the intensity, so you really want that their overlapping. So, they are really balancing, there could be a small variability in the beginning owing to the fact that they are just the background and then the spots coming in. So, 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, you can see and select whether, yes this PMT is being good for me. So, this is a way you check which PMT is more suitable. So, you select the PMT, look at them, see the overlap, wherever there is a best overlap without saturation, you want to go over those settings.

I guess, now it is time that we should really move on to the, your software interface and just see like how really this scanning is performed by using this software. Yes, let us switch on the GUI. So, Pankaj, so let us talk about the software, how actually we can use this software for image scanning from the hardware so that, now we can see it live that,

how one can actually control different buttons for acquisition of the good images. Let us go to the GUI. Sure. One needs to explore.

(Refer Slide Time: 36:32)



So, what you see now is a GUI or graphical user interface of GenePix pro software. Yeah. On top, it is in the form of different tab buttons which allows you in a different work group. See, for example, image allows you in different ways and crontoning of the image acquisition and histogram looks at, how that image has performed. So, this is what we were speaking about in the earlier slide where you can see a live kind of demo which is happening and then lab book actually gives what all you have done in a different step wise. So, what is being every movement of yours in this particular software is being logged in and analysis can be done in the form of batch form which allows multiple slides so that you can do alignment and the analysis which can be performed with the batch analysis. And once the analysis is over, results can be seen and scatter plot can be now plotted in the level of this graphical user interface; once you are through, you can look at the reports as well. So, let us look at the major function of imaging, so what and how one control for the best image acquisition?

I think that is important, let us look at some of the buttons and how one can control those acquired image. Let us quickly go through like different kind of buttons here. So, now the imaging can be done at a different wavelengths and the like, preview can be done at 635 and 532; in a single laser based, the wavelength can be done at 635 or wavelength at

532. So, even the ratio of the imaging, how the both has performed together can be looked at looking at a 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.

(( ))

And now, let us 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 chips in the form of this hand tool, the plus indicates the zoom tool and the other tools are actually this is one to unzoom and you can also look at the whole image button. So, once you have the image, these two becomes activated, so these two are actually for the blocks. And looking at the controlling the blocks and these ones are the features. So, many a times what happens, usually you get the GAL file which is actually the feature information file.

Yes, actually you can just maybe elaborate on gal file, because it is one of the very (( )) commonly used, many people talk about micro (( )). Sure, so GAL stands for gene array list. So, actually it gives the X and Y coordinates where each array usually is being presented in the form of blocks, which blocks in turn are in the form of features. So, these blocks and feature positions are being recorded in GAL file and then the information or annotation is given to each bar. So, GAL, G A L that is gene array list file, essentially contains the X Y and the number of columns. And so, also the information of each part, how they are being annotated and placed on the chip. So, if by chance if you do not know or you are prepared by yourself, these buttons here, allow to make your own blocks and create your own GAL file with the help of the tool which is called as gene array list generated. Ok.

So, now let us look at the control button which is towards your right side. So, the first one is a preview scan and then you can also have a data scan. (( )) One stands for one wavelength, so it allow you take image from only one wavelength. Sure.

And then you also have a multiple scan. So, you do a preview scan and then you do a scan with this button, you also have other buttons which will light up as and when you acquire the image and this is for the analysis. So, the once the analysis is being done, if you click this button, the analysis will be performed after the alignment. So, this is

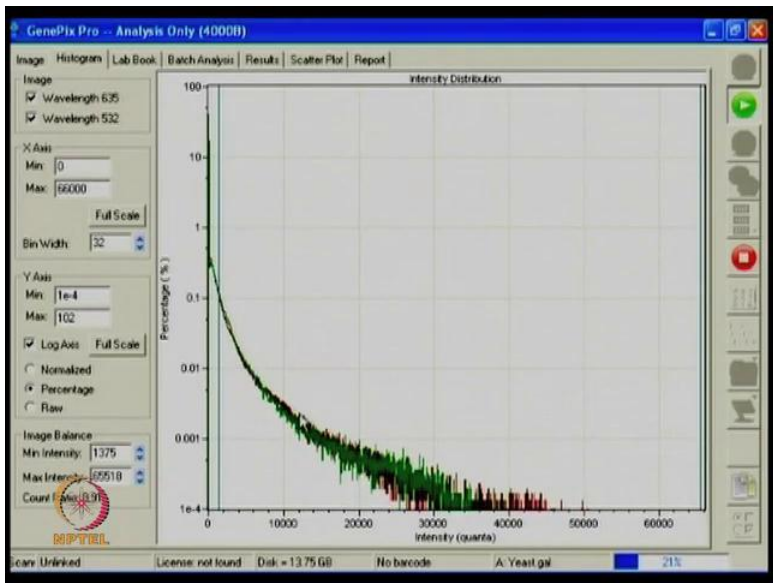
actually open button, so this is like normally your file where you want to open or save your images.

And this one is actually a flagic, as we discussed the different features can be flagged, you can look at when the image is available to you, you can look at good, bad or absent and you can give them a different ratings. Here again, he is looking at different zoom buttons, so which allows you that, which you want to focus on feature name, so the feature I D's where you want to go for one particular (( )).

The major one here is particularly this, which allows a different workflow controls right. So, this allows in the form a different ways can be had, now let us quickly go through a one particular scan which is a simultaneous simultaneous scan. So, both (( )) will be acquired at the same time. So, if I press on a data scan button, the image, the after putting a inverted slides in the hardware, it is scanning. So, you just see on the top which is less visible, so let us try to zoom inside. So, if I just put this button and allow you to zoom, (( )).

You can see, particular how the scanning is happening. So, you are looking at different image type. So, if I click on only one wavelength this, because it is live after the scanning, you can see it is going for the ratio image scan.

(Refer Slide Time: 41:54)



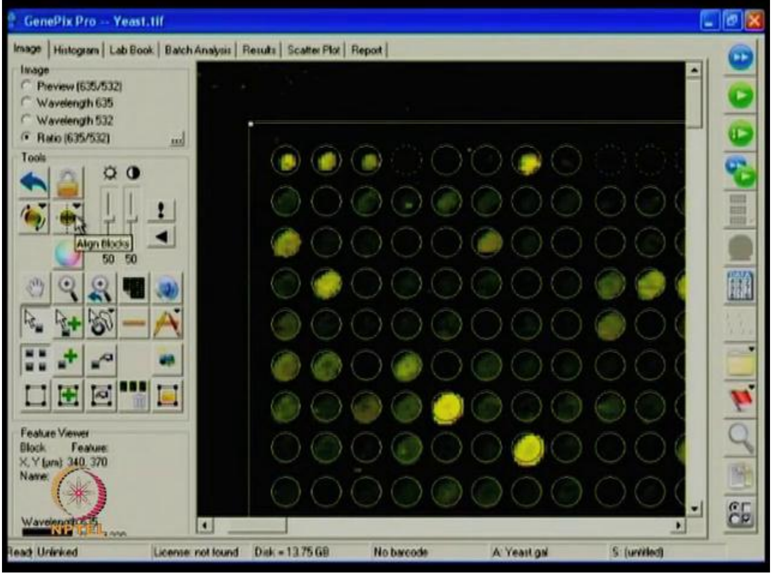
So, now quickly at a histogram, you see it is start coming up, because if this scanning is going on live. So, it is start reducing, the basically as we discussed it should be overlapping. So, my settings are usually looking very nice in this particular one. So, I guess as this scanning is progressing, one need to also keep looking in simultaneously the histogram. Correct. To determine like, how side 3 and side 4 are how well aligned.

Correct. So, how well aligned with the help of auto PMT, so side 3, side 5 you can adjust, auto PMT you can adjust, laser power so that you can see this one. So, if we see some variation, then we need to come back here and adjust this parameter so that they are super aligned. Super aligned, yes. So, in this fashion, the image acquisition is being happening. Right.

And you also have, you we give you a power that in between the slides usually people keep barcodes and our system or GenePix pro is compatible with reading the internal barcode which is being done so that you can have multiple scans also being passed by. 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 upgrade available. So, now as the scanning is being performed, let us look like, I will save the image. Right.

And once I save the image, I would like to see how the different process is being done. Yeah. So, say for example, I have saved this image in the form of say, this is the yeast. So, I just want to open an image which I have just saved.

(Refer Slide Time: 43:36)



So, basically as we discussed, each particular array can be divided into the blocks. So, this particular array of the yeast contains four blocks and each block is having features. Correct. So, number of features information is given in that, so basically 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 G A L or a GAL file which allows me for alignments.

So, best feature of GenePix pro is its capability of identifying feature by itself, it is little little (( )), so let us just see, how the zoom button looks like. And when you fine tune that alignment for overall proper image extraction So, only thing you have to do here is just take your block and allow it to move 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 use the first pitch finds all features, all blocks and do automated fashion. So, if you click once, you see software automatically finds all its features.

Wherever by chance the features are absent or there are some physical deformity, it says it is not present or you flag it as bad. So, what say if it is good here like it is automatically adjusting to the spot size. Correct. The overall width, if it is adjusting according like with that some spots are not so uniform right. Correct. So, it is making that correction here.

Correct, so as you see, you can actually move it, but it does not affect life as long as you have just kept once and the data is being stored. But usually people ask me is it good idea if I by chance move to do once again. So, it is not a bad idea to do, because it takes few seconds to do it. Right. So, once you have done this particular alignment, let us look these two slides which I said these can zoom out.

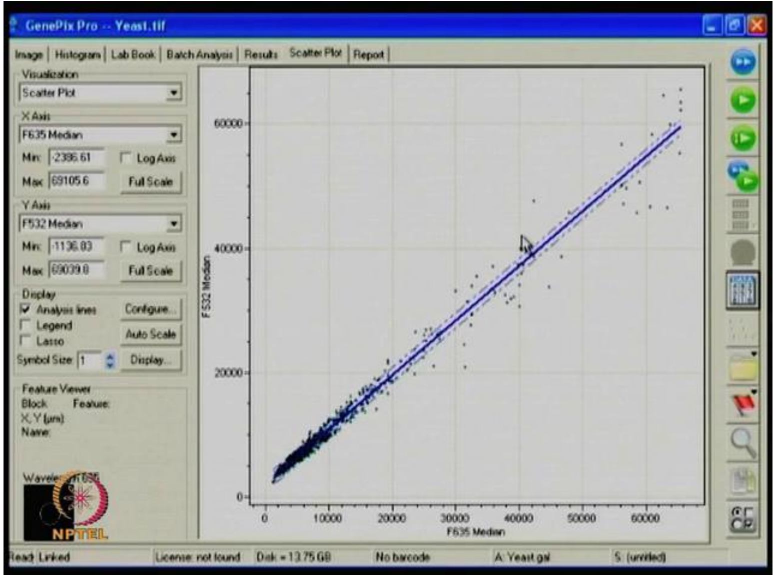
Yeah. So, you can see whole particular slide now is being scanned and aligned as well. So, it is a very quick process which software performs very easily for doing the job. And once you have done this, you can always hit a button of results, now if I go to results actually, this is empty.

Yeah. So, if I click on results, results are being calculated and there are some 40 different columns which will be output in the form which GenePix pro understands different ways. Sure. So, just quickly looking it the major ones, the major ones here are looking at this F means the intensity from different channels, 635 or 532. Right. And this background calculation is being done accordingly in the same laser range. So, once you do a corrections, what happens is you want to correct your intensity mean values with the values of the background.Ok.

So, this is what is the most important which usually people use for the further calculation. Right. Apart from a ration of means or ratio of medians which can be calculated again and being presented to you in a different column formats, so each column signifies different ones like 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 and then allowing you to align and then do the analysis. So, this is the basics steps which anybody or everybody want to do in micro steps.

So, once you see the images, the the people end up in the form of results and you have different columns available to you. Right. After that, what is next thing to look for like, how good the scatter plots are or? So, there are different ways people want to visualize how my column, because number make very less sense.

(Refer Slide Time: 47:40)



So, the best way to look is a scatter plot. Scatter plot allows you in a different ways what you are plotting at X and Y axis and here if you see, I am just plotting actually towards F 635 median over the F 635 median. So, you are comparing two different channels, how they have behaved. So, essential rule is they should mostly the microwave assumes all the six chips are having the spots and which are genes which are not varying too much. So, you expect most of them to stand nearby to the origin of the centre.

Yeah. So, this is what you want to look at. How close are they aligned? How (( )) closed are aligned. (( )) Yes, they should not be too far away from each other so that they are not (()) actually to each other. Yeah. Because you expect there is few differences, but not very very significant which can be seen at very large scales.

So, I guess we have talked so far away about, how to use hardware to scan a slide and by using this software interface which is GenePix pro here to acquire all the data. Now, next challenge is how to really obtain some meaningful information from this whole data which we have already acquired.

True, so GenePix pro as we discussed in 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 a single or multiple data to handle with.

So, Pankaj, it was good to talk to you about how one can use the DNA micro arrays or even protein micro arrays, different type of slides and use the hardware disk scanner to scan this lights by using this software which has various features and depending upon experiment, one can actually take a decision that what type of features one need to use for that scanning. I think it will be very good if you can just share with us the video which describes the overview of the whole process starting from the sample preparation followed by image acquisition as well as image analysis, so if you show us that video that will be great.

Sure, thanks to molecular devices, they have provided me with this video which actually takes you from the basic process of biology in very very brief to important software parts and so, also the hardware design which is being emphasized to the level of results. So, let us watch that video.

Molecular devices introduce the world's simplest most reliable automatic microarray slides scanner. Now, you can walk away from scanning where the GenePix autoloader 4200 AL automatically loads, scans, analyses and saves results for up to 36 slides, the autoloader accommodates microarrays on standard glass microscope slides labeled with up to (( )) and (( )). These microarrays can contain just a few hundred parts or tens of thousands of parts representing an entire genome, as many as 36 slides can be loaded into the convenience 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 savings 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 notify you remotely when your batch is complete. Using the defined scanning parameters, the precision robot arm leaps in the action and moves to the first slide, a unique never let go grippers securely clamp the slide and carry it to the scanning area, a barcode reader recodes the barcode and then the slide is positioned for scanning. The GenePix autoloader 4200 AL can be configured with up to 4 lasers, a mutual density filter wheel can be used to attenuate the laser power if necessary for a especially bright samples. The laser excitation beam is delivered to the surface as the microarrays slide, the beam scans rapidly across the short access of this slide as the robot

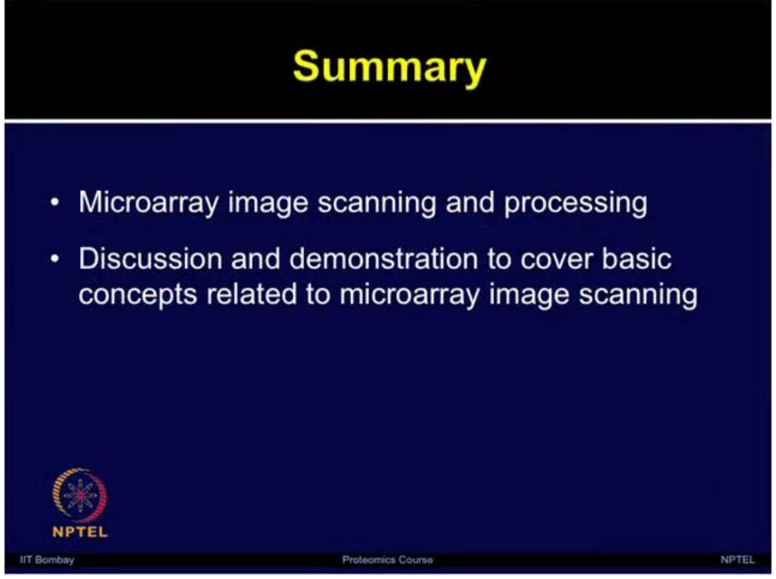
arm moves this slide more slowly than the long access, and fluorescent signal immediate from the sample is collected by a photomultiplier tube.

As this scan proceeds, sensors detect any non uniformity in this slide surface and the robotic arm adjust the slide position accordingly to ensure the arrays surface is always in perfect focus. Each channel is scanned sequentially and the developing images are displayed on the monitor, the multi channel thick 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 up the next slide.

As each slide is scanned, a list of each saved image with its associated settings and analysis files accumulates in the batch analysis tab until the batch is complete. GenePix pro automatically finds these paths and calculates up to 108 different measurements for each part, the result are saved as a GenePix results or GPR file. GPR files can be saved automatically to the acuity database for statistical analysis, clustering and other advanced investigations.

It was very useful discussion Pankaj, with you 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 we have a quite good images, then only one can do good data analysis from that. So, thank you for coming here and sharing your experience for overall microarray scanning and the data acquisition. So, thank you once again for coming here and discussing and sharing about your experience for GenePix pro, thank you. Thank you.

(Refer Slide Time: 54:31)



So, in summary, today we talked about microarray image scanning and processing, we had a discussion and demonstration to cover the basic concepts which are related to microarray image scanning and processing. One need to look at various parameters while scanning the 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 discussion on microarray work flow and how to analyze the microarray data obtained from these images which we have discussed in today's lecture, thank you.