Interactomics Basics and Applications

Prof. Sanjeeva Srivastava
Dr. Joshua LaBaer
Department of Biosciences and Bioengineering
Arizona State University, USA
Indian Institute of Technology, Bombay

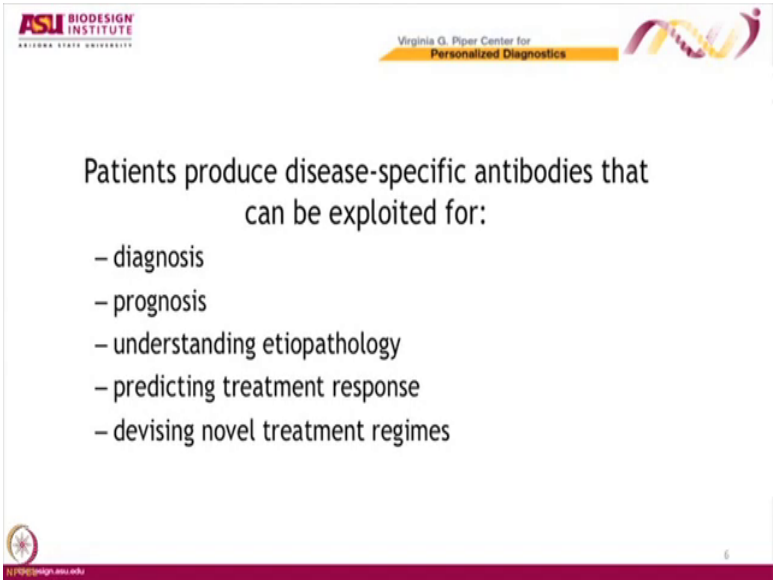
Lecture – 11
NAPPA and its Applications in study of Antibody Immune Response in Disease and in Drug Screening – I

After studying nucleic acid programmable protein arrays or NAPPA technologies and biomarkers, today professor Joshua Labaer will talk to you about how to discover biomarkers in context of breast cancer. Let me hint you the Dr. Joshua Labaer is also medical Dr. is specialized in breast cancer. So, he brings that perspective not only as a researcher but also as a clinician to give you a real good understanding about how to use the technology platforms for a very relevant biological problem breast cancer.

This lecture will be more like a case study, where Dr. Joshua Labaer will walk you through one of his approaches. And how to do immune profiling using protein arrays. So let me welcome Dr. Joshua Labaer for his lecture.

We have talked a little bit about the production of NAPPA, we have talked about the concept of NAPPA, we have talked about the concept of biomarkers. So, now, I want to talk a little bit about developing biomarkers. Specifically in the case of breast cancer in particular and how we went about that and maybe illustrate some of the things that one needs to think about in doing that. So, I think I already mentioned to you this idea that in certain diseases, individuals produce proteins that may induce an immune response particularly a b cell.

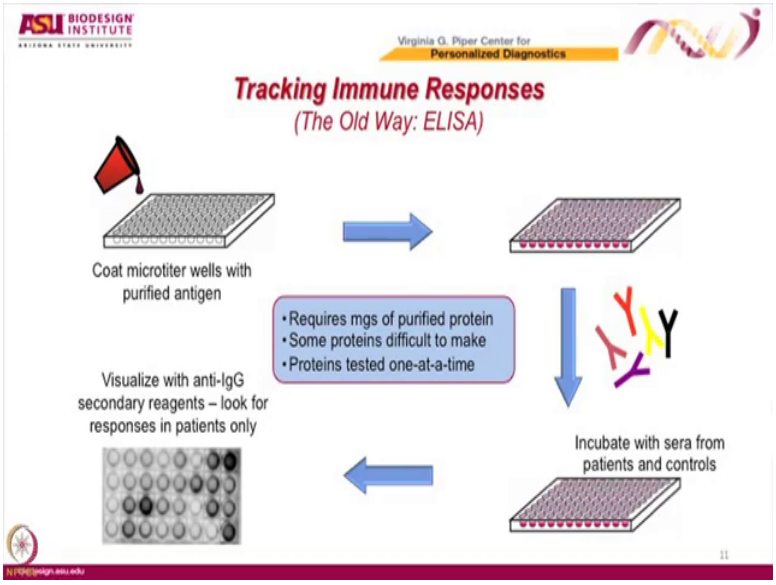
(Refer Slide Time: 01:55)



Immune response that leads to antibodies and that those antibodies can act as biomarkers for disease. And so, we have talked today a little bit about using biomarkers for diagnosis for prognosis. In this case this is not so much for biomarker use, but the fact that these are aberrant proteins in the disease might shed some light on the mechanisms of the disease.

So, the fact that the body responds to these abnormal proteins might be telling us something that is important. And so, you might look into this to understand the disease itself. Hopefully you could use this to predict a treatment response in some cases or even help us develop novel treatment regimens and that becomes especially true, if you believe the possibility of using these antigens to vaccinate the patient against the cancer. Can you induce a stronger immune response to kill the cancer?

(Refer Slide Time: 02:53)

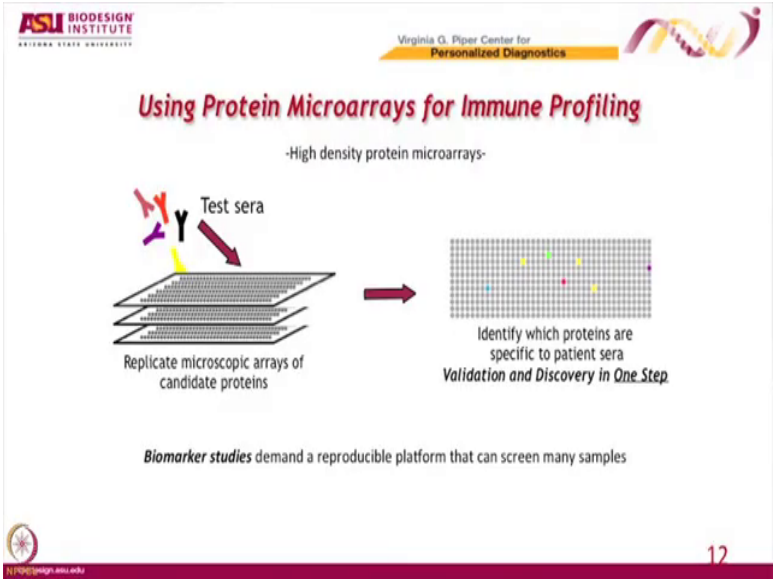


So, you remember I think we talked about this but I am going to reiterate a few things in the next few slides. That the classic way to measure an antibody response is this essay called the ELISA assay; which stands for Enzyme Linked Immunosorbent Assay but nobody says that they just say ELISA. And the idea is you put your protein in the wells of a dish, you attach the protein to the bottom of the well, you then add to the each what each well the serum from a patient; if the if the patient has a strong response to that protein then you will get a strong response like this. If the serum has no response you will get no signal there ok.

And so that tells you that each well tells you which patients had a response or not. The challenge of course, as I mentioned before is that it requires a lot of protein to coat these wells some proteins are not easy to make. And you are testing proteins by this method one protein at a time. And if you want to think like a modern systems biologist and you want to think at

scale, you would rather be testing thousands of proteins at a time. And so, that was this and then this time I have the picture there so its much better.

(Refer Slide Time: 04:12)

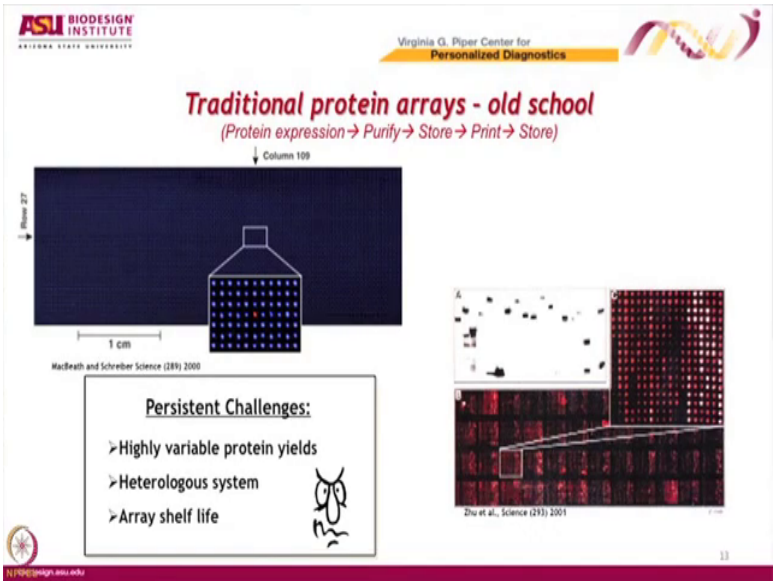


That is that is the idea of these arrays you take an array that has a couple thousand proteins on it you add patients serum and then certain spots on the array light up. And looking at the size looking at the intensity of that response might give you a clue as to how strong that patient responded to that antigen but, you also get essentially a listing of all the different proteins that the patient blood recognizes.

For this method to work you need to know that this protein array platform is reliable; you need to know that its that that when you run it on different days, using the same sample that you are going to get the same answer. Because if your gone its one thing to do research and just hope that it worked pretty well and get some responses that you can then follow up in

other studies. In this case you are going to base your clinical decision on whether or not this is a predictable marker and that only works if the platform is reliable ok.

(Refer Slide Time: 05:13)



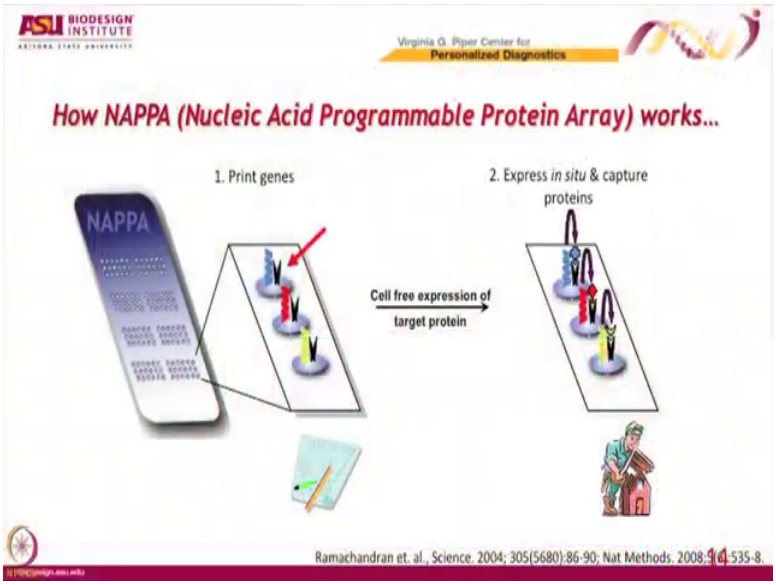
So, I think I went through this a little bit before, this is the classic way of making arrays which is to purify the proteins and spot them on the surface right. So, you take purified proteins and print them, now its a method that that definitely works people like Hong Joo have done this for many years; they purify proteins in high throughput it has limitations. The amount of protein that you end up printing varies over several logs so, by several orders of magnitude.

So, much of the protein tends to be on the lower end of that spectrum. So, very weak amounts of protein added and then a few proteins maybe you know hunt 10000 times more than that. And so, you end up with an array that might have this look where you see some areas of

strong signal, but then lots of areas, where its relatively black because there is almost no protein there.

And of course, if you are doing a biomarker study and what your intent is to look at the signal of specific spots and there is very limited protein on those spots, then you will not know if the limit of the signal is because there was no protein there or if because its a weak interaction. And so, that that is one of the challenges.

(Refer Slide Time: 06:26)



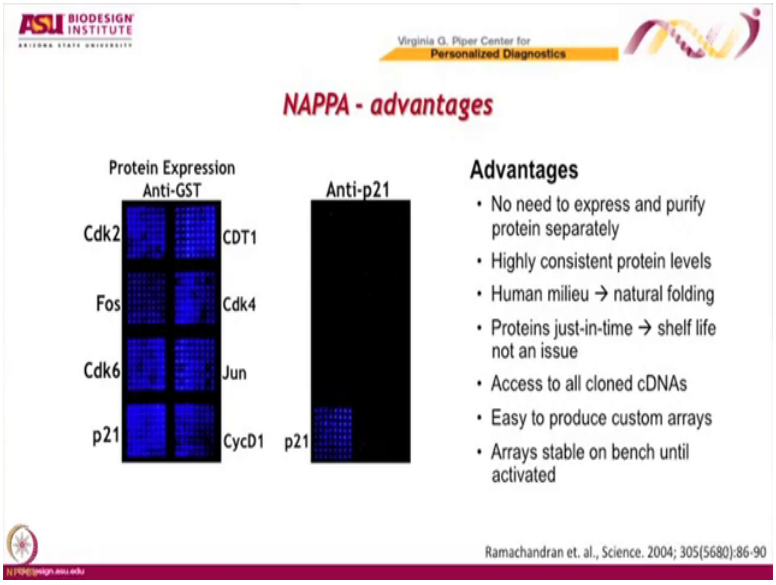
So, the approach that we came up with is this nucleic acid programmable protein array, where we print the gene for the protein on the chip. And then we synthesize the protein insights you capture to the spots and then display the protein after its been captured. And of course, we store the arrays in this state here when they are unexpressed. So, they are just DNA arrays and they are very stable.

And then once we make the protein here as you guys are doing in your experiments within hours we immediately probe it with another protein and test the fresh protein with possible interactors.

(Refer Slide Time: 07:00)

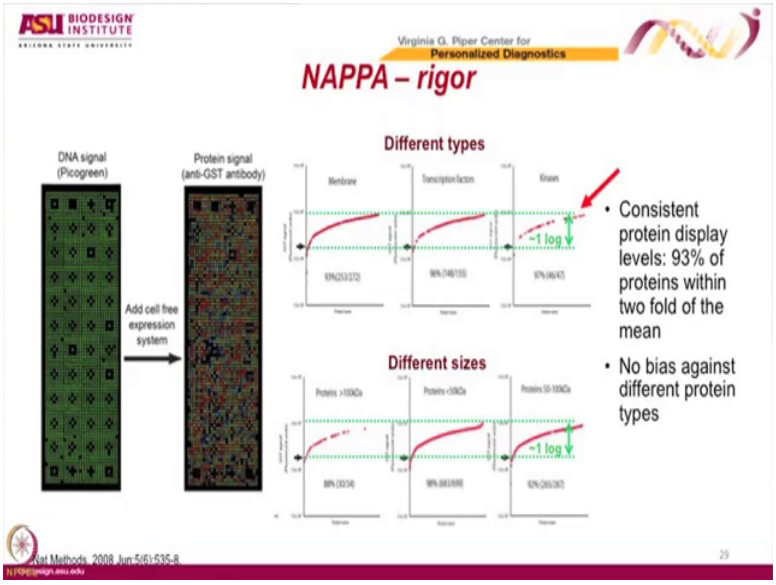
This is the repository of clones that we have been making and I think I showed you this collection here this is the actual freezer that we have and I went through all these various characteristics here. So, this is the large collection.

(Refer Slide Time: 07:14)



Of course, I also showed you this is taking NAPPA array and we test it for total protein levels by looking at GST; because all the proteins on the array have a GST tag on their c terminus. So, if we get staining of GST it tells us how much protein is present. And if we stay in the array for one protein we get that one protein and I think we have been through all these sort of advantages of this approach ok.

(Refer Slide Time: 07:39)



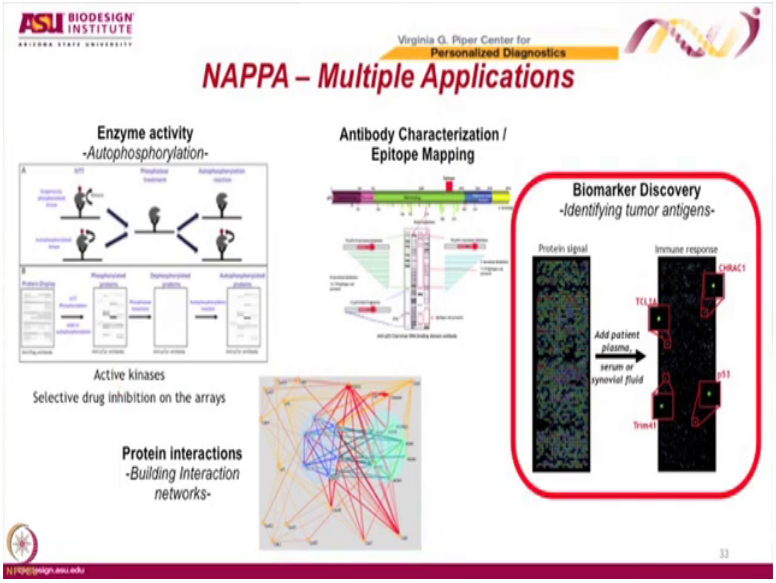
So, let us talk now a little bit about rigor. How do we know that, what we are looking at is good? Well, the first thing we want to measure is how well does the platform express all proteins right. So, we are going to take the array we are going to test it for DNA binding which tells us how evenly we printed. And we are going to test it for a protein expression which shows us that we are displaying all the different proteins and we are displaying them at an at an reasonable level.

And then this is what that looks like its summarized quickly but it shows you four membrane proteins transcription factors kinases small large medium and small proteins. The green line on the bottom is the level of detection. So, if you are above that dotted green line it means you can detect it; that means, its like 5 standard deviations above background.

And then the top green line of course, is maximum detection. And what you see is almost all the proteins fall between these two ranges and its only a single log. So, no protein is present that is more than 10 times than another and the vast majority of proteins are within two fold of the means.

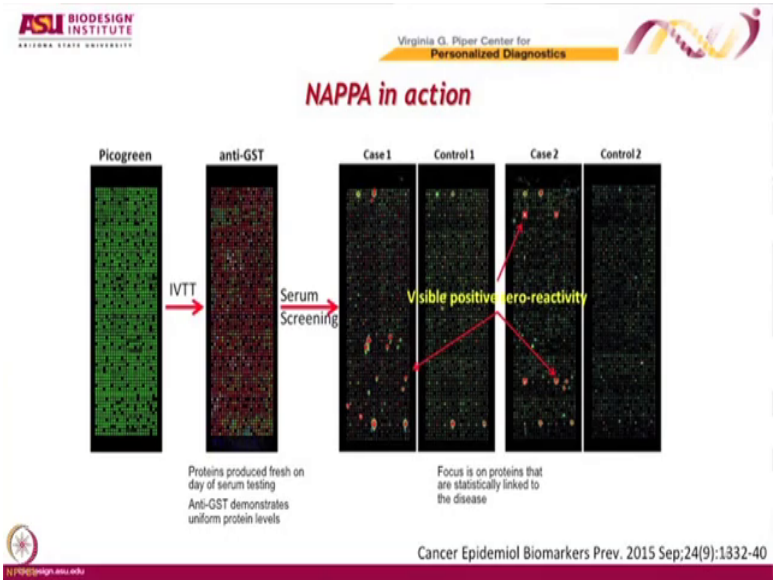
So, they are all very close together Ah right I showed I showed you that. So, its key if you are going to do a clinical study to know that there are no biases against specific protein types. If there were biases, that would be a problem when you try to do conclusions. So, these are some of the things that we have done with it about the protein phosphorylation and I probably do that in my last talk ah.

(Refer Slide Time: 09:09)



I talked a little bit before about mapping protein domains, we talked about the protein interaction studies. And now I am going to focus a little bit more on biomarker discovery alright. So, this is what we are looking for in biomarker discovery.

(Refer Slide Time: 09:33)



So, here is the DNA array, here is after making protein and here is after adding serum. Wherever you see a bright spot especially a red spot; that means, that the patient is making very strong antibodies to that protein ok; the color by the way is a false color. So, these are not its not really red what we do is, we get we get a readout of signal intensity and numerical readout of signal intensity and then the software adjusts the image and the color that it represents tells you which what level of intensity that equals.

If we were to do just grayscale intensity then we would have little pretty much only 10 folding range by doing different colors, we can cover a larger range. And so, typically blue green is

blue is weak, green is stronger, orange is stronger than that, well yellow is stronger than that, red is stronger that, orange stronger than that and red is the strongest all right. So, one thing that you notice when you look at this is that is that we have very uniform levels of protein.

So, these are all roughly the same intensity there are some a couple weak spots here, but for the most part they are all very strongly expressed. And of course, I already mentioned that we have expressed them freshly. And then the key here is to compare cases to controls and look for proteins that are only present in the cases. So, that is first part.

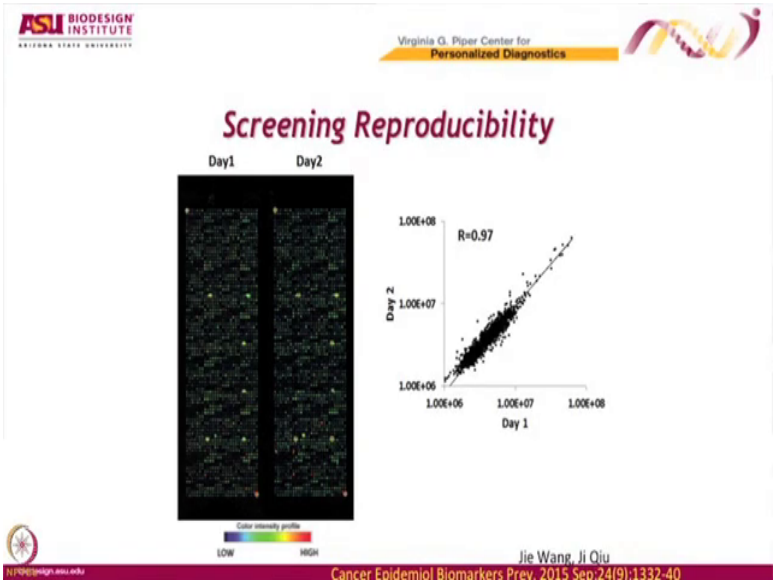
So, now, we know; we know we talked about the advantages of making fresh proteins getting different classes of proteins well expressed. And looking at the proteins, in you know that being folded by natural chapter 1 proteins and a natural lysate. Now, what we need to know is, if I do this experiment today and I do it next week will I get the same answer.

And so, the way we went about testing that was we created a control sample. And to make a control sample we took several sera from different individuals; mix them together to make a large volume of a mixed sample that would have a lot of responses now it has a lot of responses because we mixed it. And because we now have a large volume we can use a little bit of that sample every day we do the experiment.

So, every day we do the experiment we take a little bit more of that control sample and we get the data for that sample and then we can compare that sample to yesterday and to the day before yesterday and we can ask do we get the same signal every day right. And so, I recommend doing that if you are going to do a clinical study to build a control sample and run that sample every day you do the study. So, you can say that day everything was working well.

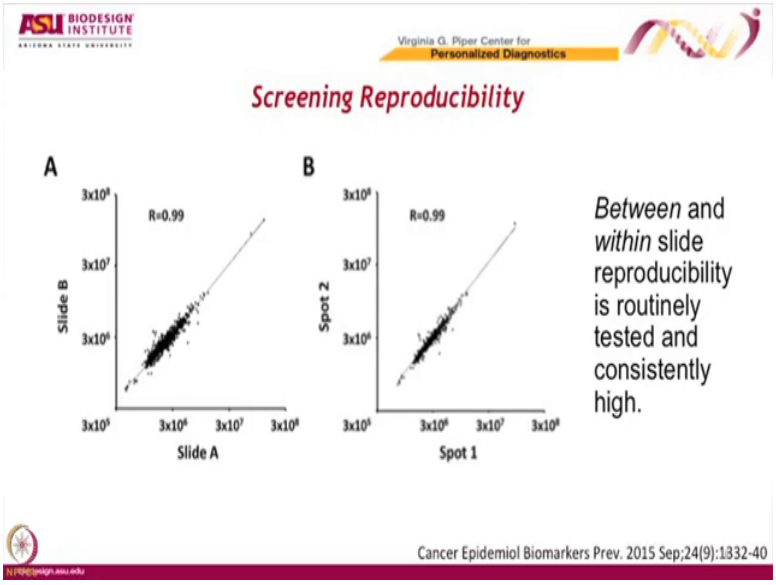
If we see that the control sample deviates significantly from, what it looked like on the previous days will throw out the data from that day; its just not worth it you know I would rather have clean data. And so, this is what that looks like.

(Refer Slide Time: 12:43)



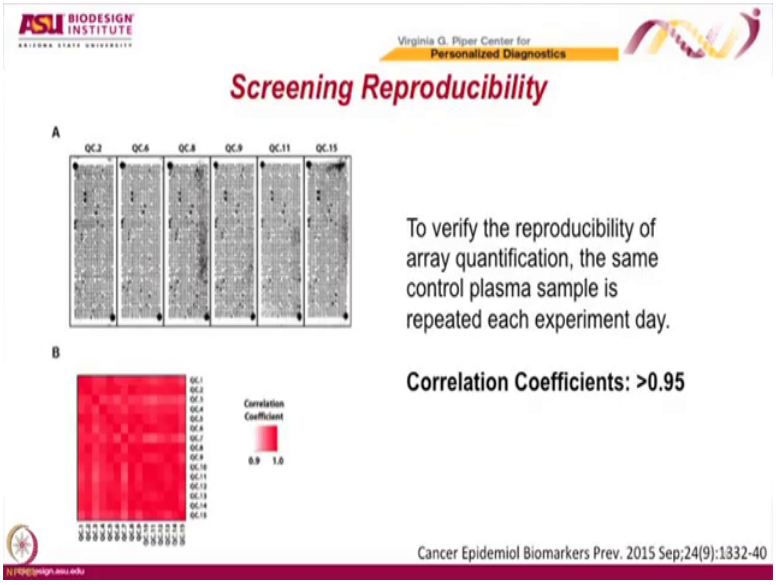
Here you see comparing day 1 to day 2; I think you can see that these two arrays look very similar in their in their intensity. But more importantly if you map all the signal intensities from all these spots a on dot plot what you will see is they all line up on the 45 degree angle here right. There the signal intensity from this day and this day for that spot is the same and so on and so forth; they line up pretty well along that line. So, that is just comparing two days.

(Refer Slide Time: 13:17)



Now, imagine if here is a couple of other examples slide A to slide B; here is another one we also often will print two spots for the same protein on the same array and the advantage of that is that we can compare intra array do we get seeing some signals. And again you see very very good correlation from spot to spot within an array. So, we have good spot we have good signal intensity within the slide and we have good correlation from slide to slide ok.

(Refer Slide Time: 13:50)



Let us look at it for the entire experiment and that is what is shown here. What we did was I told you we took that control sample and we ran it every day we did our experiment. And then here is what here is that same array on different days using that same sample and then you can plot every day versus every other day using a heat map like this.

So, the heat map is set up so, that if it were if we were bone white that would be a 90 percent correlation; and if it was solid red that would be a 100 percent correlation. So, as you can see no matter which two days we compare, nothing gets even close to white almost everything is you know medium pink all the way up to red.

And so what that tells us is, that no matter what day we compare to any other day its better than 96 percent correlation. So, everything agrees really well. So, that gives you confidence that that when you use this platform and you measure samples you are going to get the same

answer every time all right. So, then that one of the things that we want to do is, we want to be able to validate the results we get from our discovery study.

So, we talked about this a little bit earlier today, the discovery study the observed difference right that we talked about. Typically you know in the early stages of a biomarker study, you are going to begin by looking at a large number of possible variables; you are going to look at as much of the proteome as you can get your hands on right.

So, that leads you to this issue of potential over fitting and it but it gives you the advantage of having searched as much possible experimental space as possible to find any marker that is going to correlate with you know that is going to be predicted for your disease. So, when we begin these studies, we are going to study these days in my lab maybe 15000 different proteins that is a pretty good list.

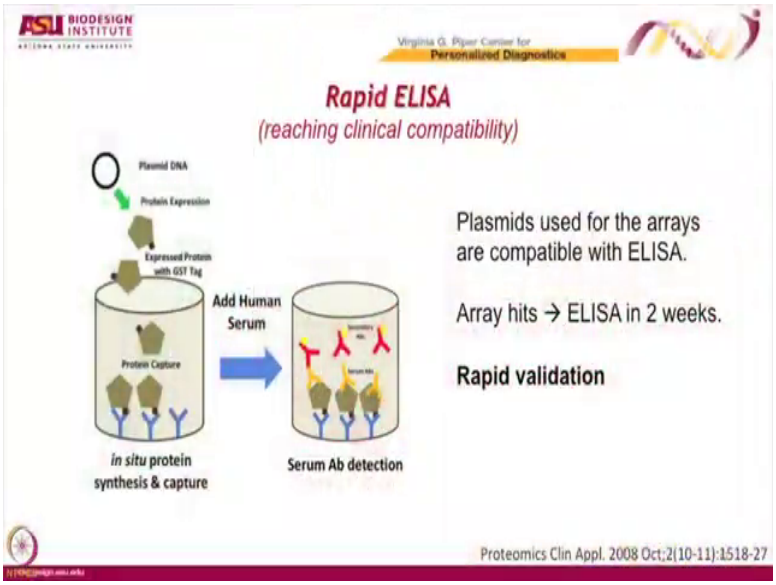
Now, I attached to that list is a price tag right, because by definition even using NAPPA to get 15000 proteins on a slide; we have to do 15000 DNA mini preps and that is a lot of DNA mini preps. It's certainly easier to do mini preps and it is to make protein, but its still a lot of mini preps. And even if; even if the mini preps were two or three American dollars per mini prep to do that many many preps is you know nearly 50000 dollars right. Just to do that is that number, not to mention the labor and all the time involved in preparing it.

So, once we have; once we have done our initial study where we compared let us say 50 cases to 50 controls right. So, that is a 100 array sets a 15000 each; when we get to the next stage we do not necessarily want to test all 15000 proteins; we have now eliminated most of those proteins, we now know that of that 15000 14850 of them are probably not good markers we can toss those out. So, now, I do not want to have to use my big chip that does all those proteins. I want to focus on 150 candidate markers in my next study and so, that is where this next platform becomes very helpful.

Now, we can come back to the ELISA assay right. And so, the ELISA assay says ok, I can make individual proteins at a time much less costly than making 15000 protein arrays right; but I cannot but I have to do them in a larger scale. And so, one of the things that we

developed in our laboratory was a way to make ELISA without having to do a whole protein purification from bacteria ok.

(Refer Slide Time: 17:58)



And we call it rapid ELISA and the idea is that we follow kind of a similar chemistry to the one that we use for NAPPA. So, we put the plasmid into the well we then the at the bottom of the well we have a capture agent an antibody that recognizes GST. We express the protein in the well we capture the protein to the GST, we wash it away and now we are left with a protein displayed in the well.

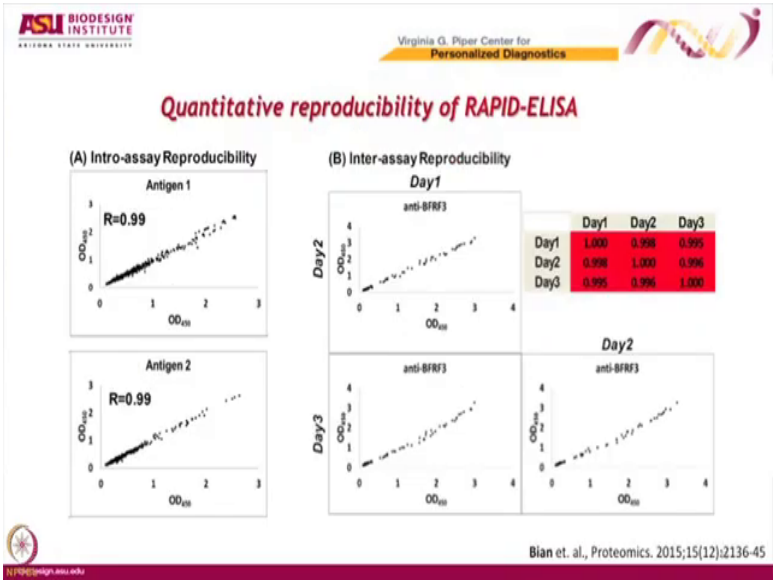
So, if you remember from my second talk where I showed you those 96 well plates the sort of early NAPPA; that is what we are doing here but in a more routine way. And we can do this to the point where I think it costs less than a dollar per well to do the assay. So, that you know

in the big picture that is pretty good; especially because we can set up one of these ELISA's for just about any protein within a couple of weeks.

One reason we can do that is by definition, if the protein was detected on a NAPPA array, then we know that we have the plasmid for that protein; because we had to print it on the array. And the very same plasmid that we printed on the array we can use for the ELISA. So, the system is immediately compatible with moving from the array to the ELISA assay. So, we can very quickly set the ELISA up.

So, we add; we add the expression plasmid we make the protein we add to the well we capture wash it away everything we do not need and then we come in with serum to the well and look to see if we get a response.

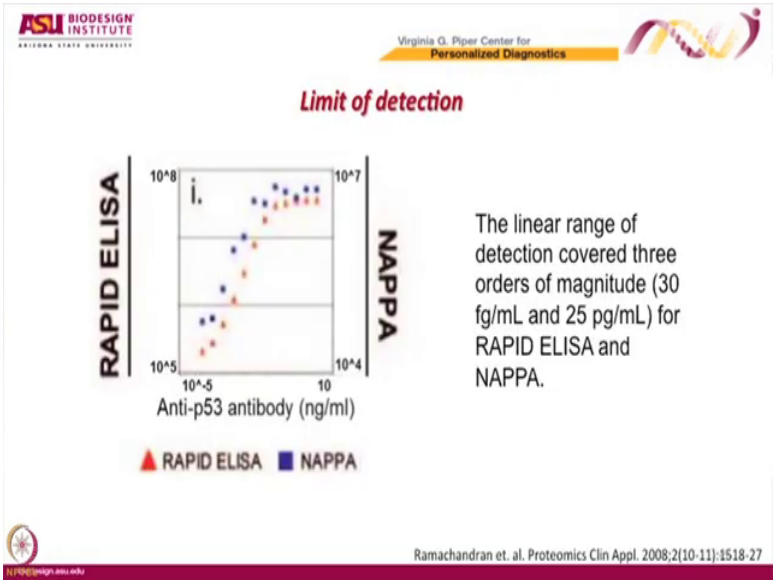
(Refer Slide Time: 19:34)



And now I want to just show you that that the ELISA assay is also reproducible. And again you do not want to trust any clinical studies where you cannot show reproducibility. So, here what we are doing is basically looking at a variety of different antigens, these are two different antigens comparing them on two different days using the same assay; using all of these different samples probably 96 different samples here. And again you see that from day to day you get the same answer when you when you use the same platform.

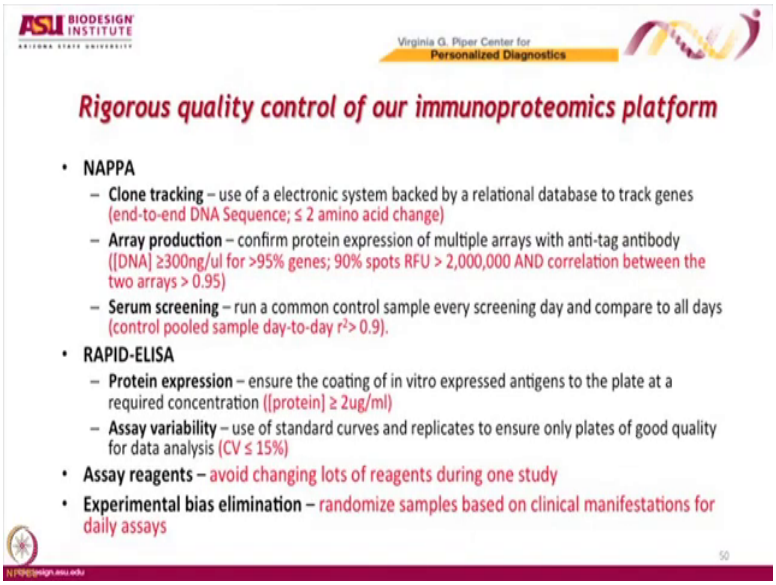
This is within an assay and this is between two different, this is within a day on the on two identical plates and this is between two different days. And then just showing you some examples and showing you that the correlations are typically close to one point zero means that they align very nicely.

(Refer Slide Time: 20:24)



We have also spent a little bit of time looking at our detection limit that is what is shown here using a purified amount of anti p 53 antibody; the assay is linear from 30 to 30 degrees to 25 pico grams per mil. So, that gives you some sense this is the range where we can operate very sensitive insensitively alright.

(Refer Slide Time: 20:47)



So, then once you go through all of this, one of the things that you have to do when you develop a platform to start studying these things is you have to start thinking about. What are my quality control checks, how am I going to make sure that as I do my experiments everything is working the way it should be working. And in particular if its not working that way then you have to you have to jettison that step and go back and fix it before you move on right.

And so, a good a good biomarker study involves a lot of q c. And so, we have a we have criteria at every step in our flow path that we follow and make a decide whether to move forward. So, I already mentioned to you before that we do clone tracking electronically we do end to end DNA sequencing. So, we have to make sure that the gene is correct by sequencing it from one end to the other. And I think I mentioned that it has to have less than 2 amino acids different. So, we will accept up to 1 amino acid substitution before we move on.

When we make arrays we do a number of things. First of all we make sure that there is greater than 300 anagrams per micro liter of DNA for 95 percent of the genes that we are going to print. So, before we even print we look at the wet the plate that has the DNA in it and we make sure that we have an adequate amount of DNA for every protein. If we do not we go back and we fix the ones that are broken or that are too low. Then when we print we express the proteins and use an antibody using anti GST antibody to make sure that there is protein present.

And we ensure that 90 percent of our spots have more than 2 million are relative freak intent fluorescence units on it. So, this tells you that we have good protein levels and then we look at two different arrays from a printing batch and we assure ourselves that they agree with each other by better than 95 percent in a correlation coefficient.

So, we are checking to make sure that we are adding enough DNA, that we are making enough protein and that the features on the arrays from (Refer Time: 22:50) to another agree with each other. So, we do that for every experiment we do ok. And then when we do? Serum screening we run the common control sample I mentioned you do that already and we show that from day to day the correlation is better than 90 percent.

If we are doing a rapid ELISA, then we make sure that we have at least that amount of protein concentration when we make them proteins. And then we make sure that the assay variability has a CV of less than 15 percent and then these other things are basic routines. One of the things I will caution you about is this little subtle point down here, that a lot of people forget

to and it can really come back and bite you if you forget to do it, which is to randomize your samples when you do your assays.

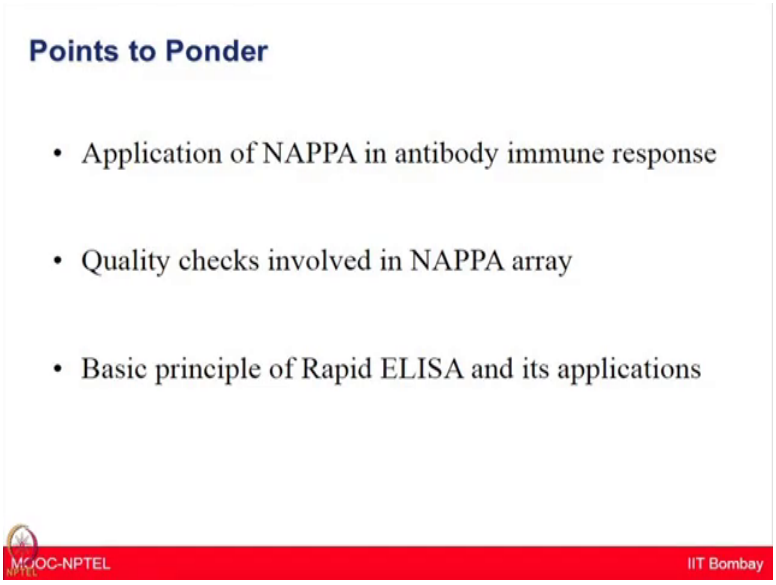
I cannot tell you how many students have had come to the lab and they will do all of the cases today and they do all of the controls tomorrow they think alright. I did it right its like no because they will get big differences. And I say I found something really cool and it turns out the difference is because they did all of the cases on one day and all the controls on the other day. And there are subtle variations to this maybe you do everything the same day, but you first load the cases first and then you load the control second. And so, the first ones that run through the machine are the cases and then the next ones that run through machines are the controls.

So, you have to make a concerted effort to make sure that you mix up everything that there is an even distribution of cases and controls in every step you do. And that there is no order bias or more plate bias or day bias to the cases or controls or else you will end up chasing your tail you will; you will think you found something really good. And then you will discover a year later after investing all of your time in it then it was an error because you just did not load them in the right order.

So, do not miss that I had a post doc who chased his tail for 4 months; thinking that he had found something really cool and when we got down to it was because he ran the case plate first and then he ran the control plate second. And the control plate sat for 10 minutes, while the case plate was getting read and that was the difference so, you cannot do that.

.

(Refer Slide Time: 25:12)



So, I am sure after attending to the lecture, you found this very interesting. You have seen how auto antibody responses for the patients could be measured using protein microarray based platform especially NAPPA technology; which could be useful for early detection of breast cancer. The intensity of signals they show how is strongly a patient response to a particular antigen. In a single experiment you will also know how many antigens a particular patient recognizes.

You now know the importance of testing the reproducibility in microarray experiments; within the slides different batches of slides, as well as your day to day variations assay. All these performance have to be recorded compared to test out the reproducibility of your data and your experiments.

Protein microarray platform is definitely very robust technology but you are say has to be reliable. And you need to document the quality control checks and the data to provide the significance to provide the confidence to the reviewers and to the clinicians who want to take your lead forward for the patient care or actual biological applications.

You studied about the challenges for encounter, while developing a biomarker and how to find solutions of these problems. Finally, today you learned about rapid ELISA; a routine technique which is developed to measure the patient response to any antigen, but at very low cost usually less than a dollar per well. This topic and lecture by Dr. Joshua Labaer will be continued in the next lecture.

Thank you.