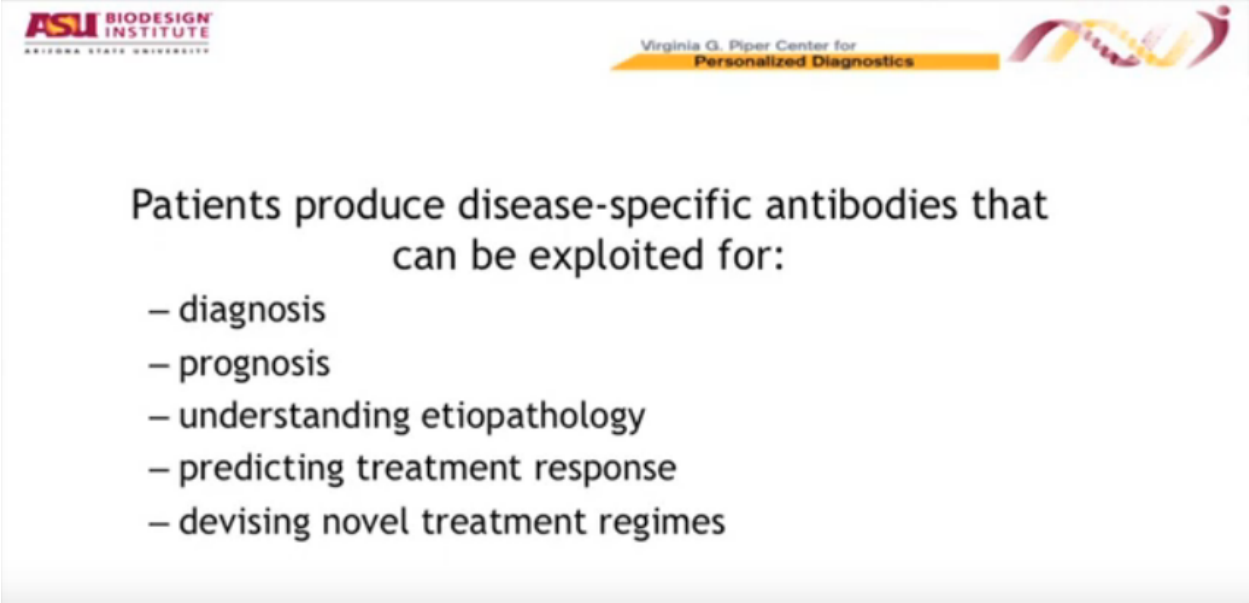
Lecture 7:
NAPPA and its applications in study of antibody immune response in disease and in drug Screening-I

Welcome, to more coats on applications of Interactomics. Using genomics and proteomics technologies, after studying nucleic acid, programmable protein arrays or NAPPA technologies and biomarkers. Today personally Dr. Josh La Baer, where will talk to you, about, how to discover biomarkers, in context of breast cancer. Let me hint you that, Dr. Josh La Baer, is also medical doctor, 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 history, where Dr. Josh La Baer, will walk you through, one of his approaches and how to do, immune profiling using protein Ares slippery welcome Dr. Josh La Baer, for his lecture. We've talked a little bit about the production of NAPPA, we've talked about the

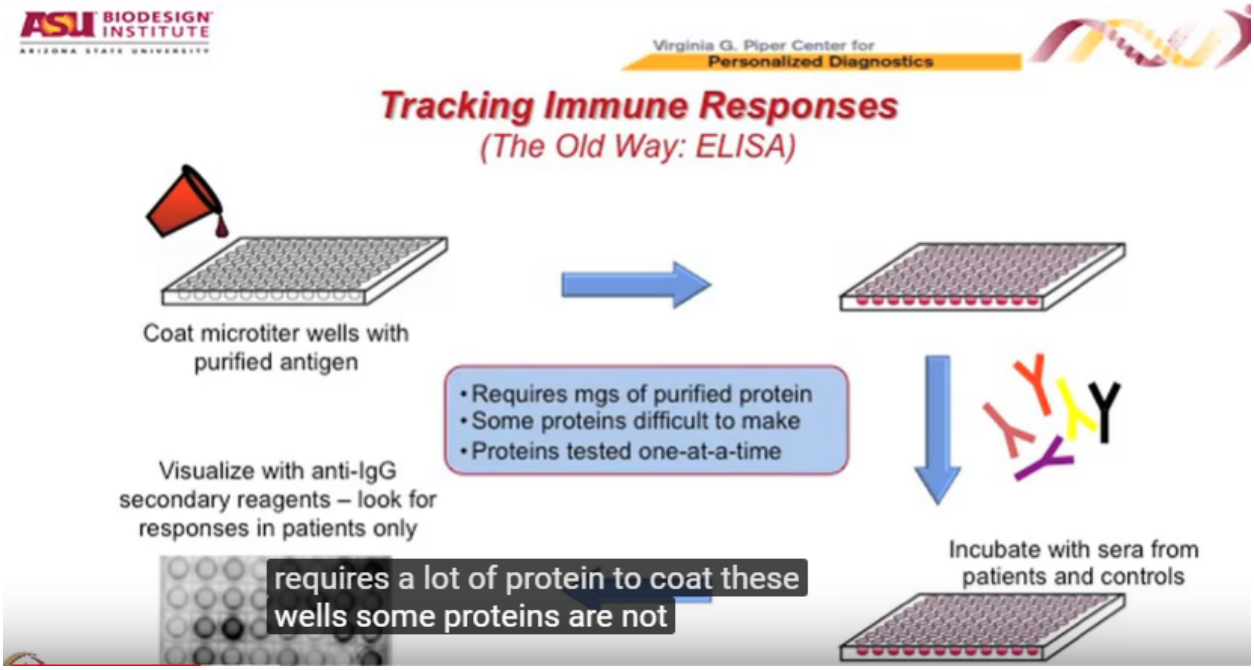
concept of NAPPA, we've talked about the concept of biomarkers. So, now I want to talk a little bit about developing, biomarkers specifically in the case ,of breast cancer in particular and how, we went about that and maybe illustrate some of the things that one needs to think about in doing that. So,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

Refer Slide Time :(2: 17)



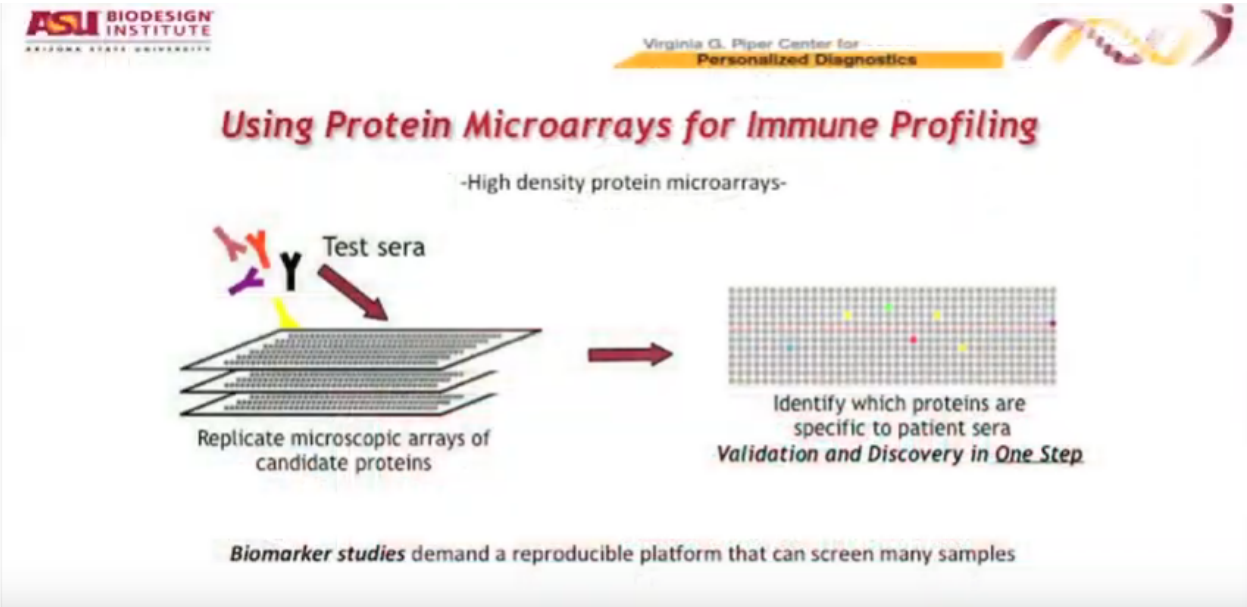
Particularly a b-cell immune response that leads to antibodies and that those antibodies, can act as biomarkers for disease. And so, we've talked today a little bit about, using biomarkers for diagnosis, for prognosis, in this case this is not, so much for biomarker use but, the fact that these are aberrant 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's 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 :(3: 16)



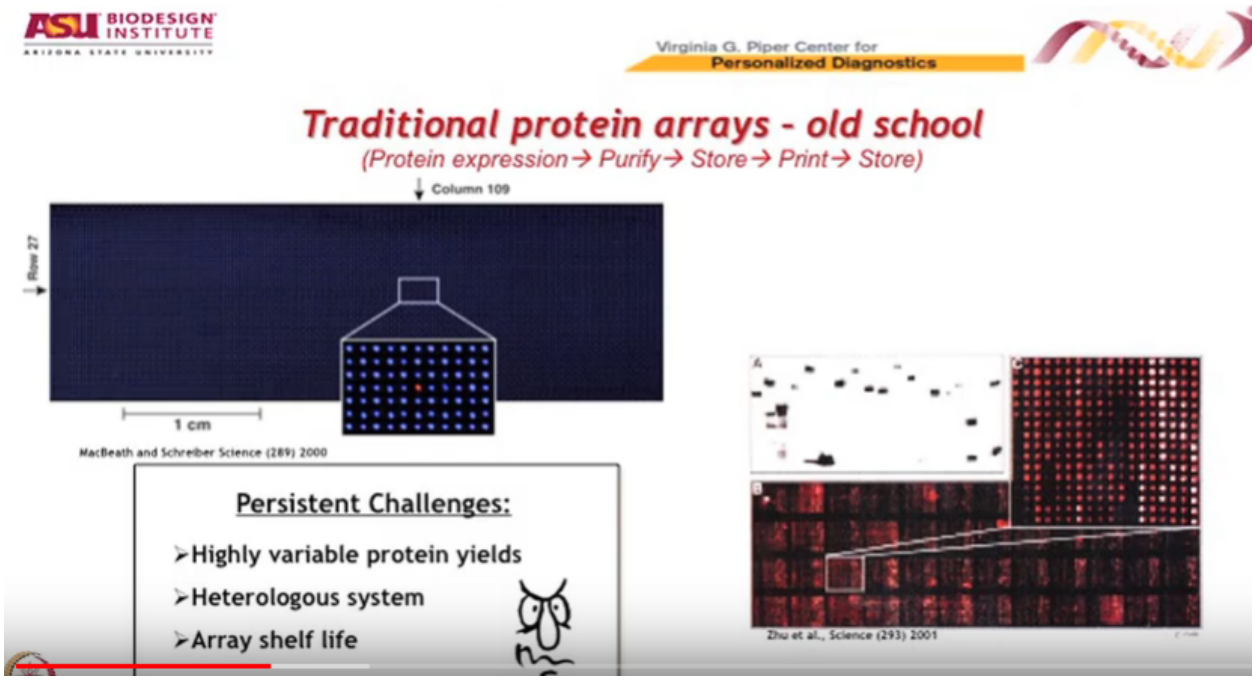
So, you, you remember, I think we talked about this but I'm gonna reiterate a few things the next few slides, that, that the classic way to measure an antibody response, is this, is this assay called the, 'Eliza Assay' which stands for enzyme linked immunosorbent assay but, nobody says that they just say Eliza 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 cirrie if the patient has a strong response to that, protein then you'll get a strong response like this, if the serum has no response you'll get no, no, no signal there. Okay? 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're testing proteins by this method one protein at a time and if you, you want to think like a modern systems biologist and you want to think at scale, you'd rather be testing thousands of proteins at a time. And so, that was this and then this time I have the picture there.

Refer Slide Time :(4: 35)



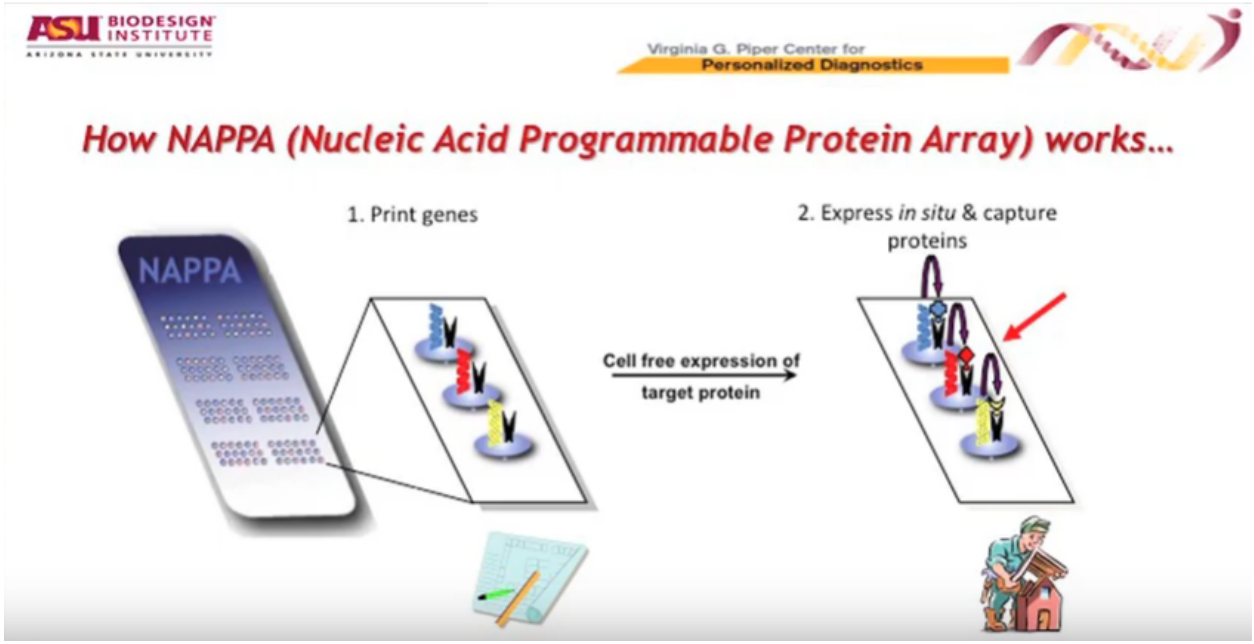
So, it's much better that's, that's the idea of these arrays, you take an array that has a couple thousand proteins on it you had patients serum and then certain spots on the array light up and looking at the size looking at the intensity of that response, might give you a clue as to how strong that 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 it's, that, that when you run it on different days using the same, sample that you're going to get the same answer, because if your gun it's 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're gonna base your clinical decision, on whether or not this is a predictable marker and that only works if the platform is reliable.

Refer Slide Time :(5: 36)



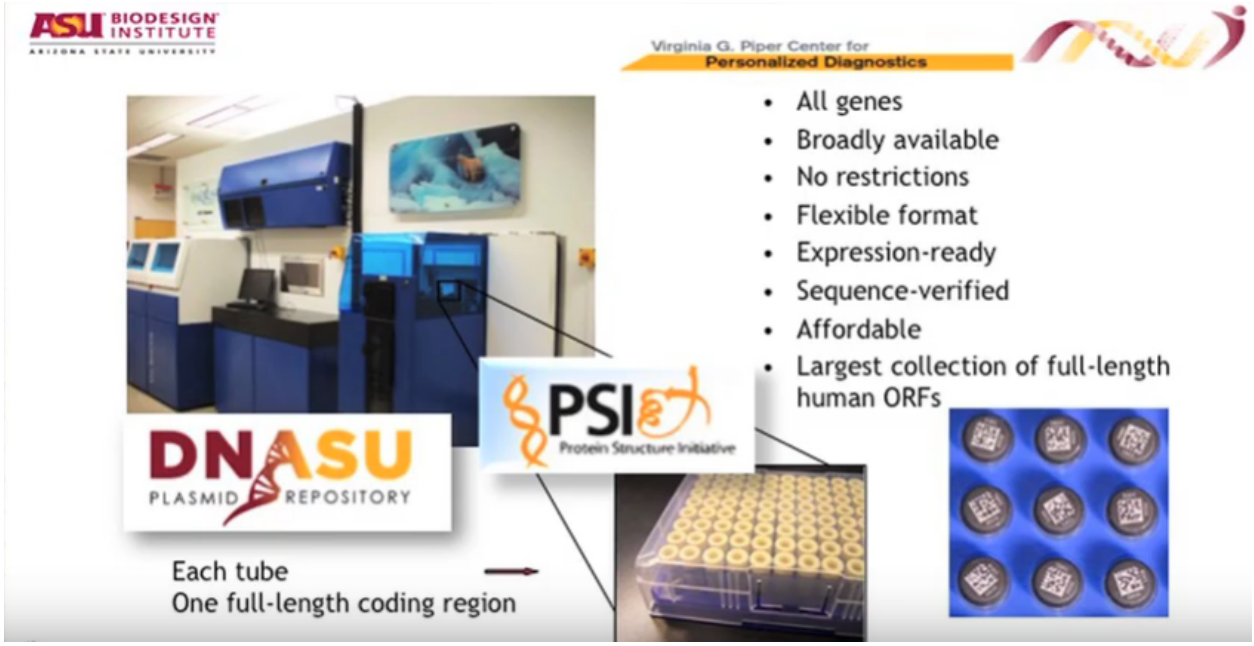
Okay? So, I think I went through this a little bit before, this is the classic way of making arrays, which is to purify the proteins and spot them on the surface. Right? So, you take purified proteins and print them, it's a method that, that definitely works people like Hong Joo, have done this for many 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 a hunt 10,000, 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 it's relatively black, because there's almost no protein there and of course if you're doing a biomarker study and what your intent is to look at, the signal of specific spots and there's very limited protein on those spots then you won't know, if the limit of the signal is because there was no protein there or if because, it's a weak interaction. And so, that that's one of the challenges.

Refer Slide Time :(6: 48)



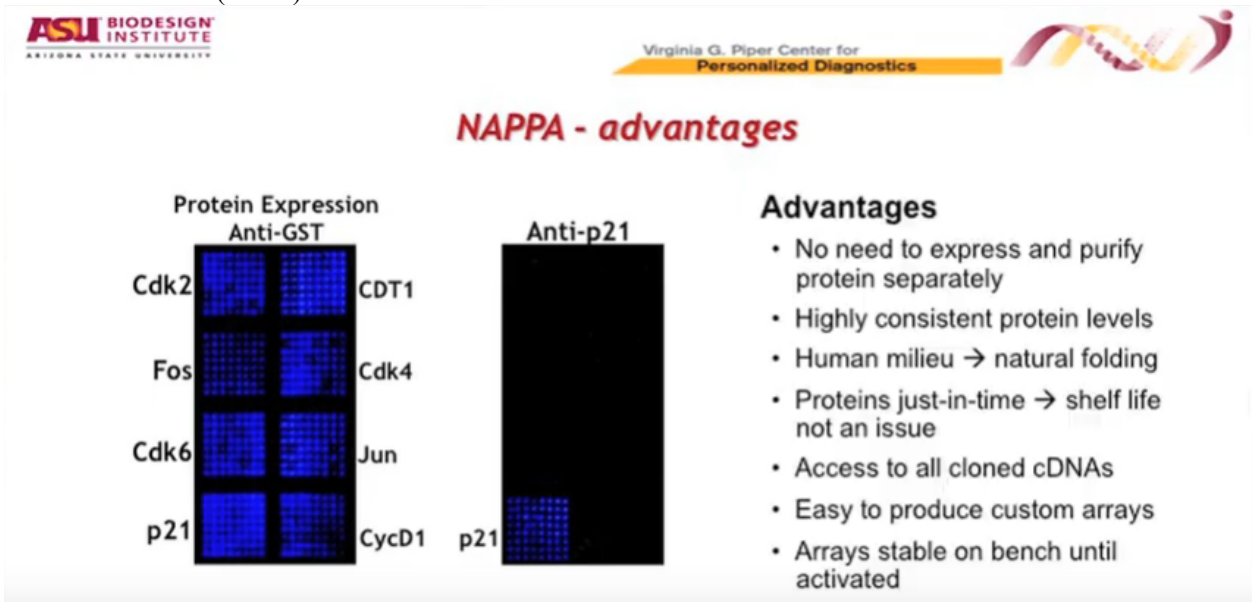
So, the approach that we came up with is this nucleic acid programmable protein array. Where we print the gene for the protein on the chip and then we synthesize, the protein insights you capture to the spots and then display the protein after it's been captured and of course we store the arrays, in this state here, when they're unexpressed. So, they're just DNA arrays and they're 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 :(7: 23)



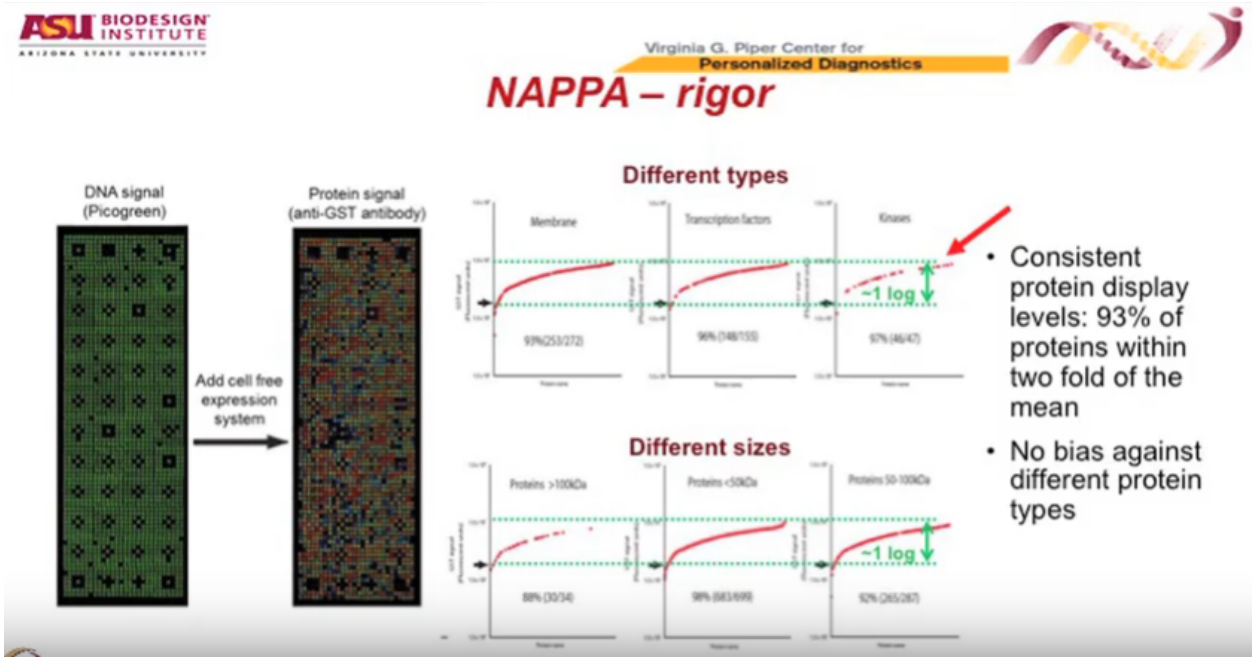
This is the repository of clones that we've been making and I think I showed you, this collection here this is the actual freezer that we have and I went through all these various characteristics here. So, this is the large collection of course.

Refer Slide Time :(7: 38)



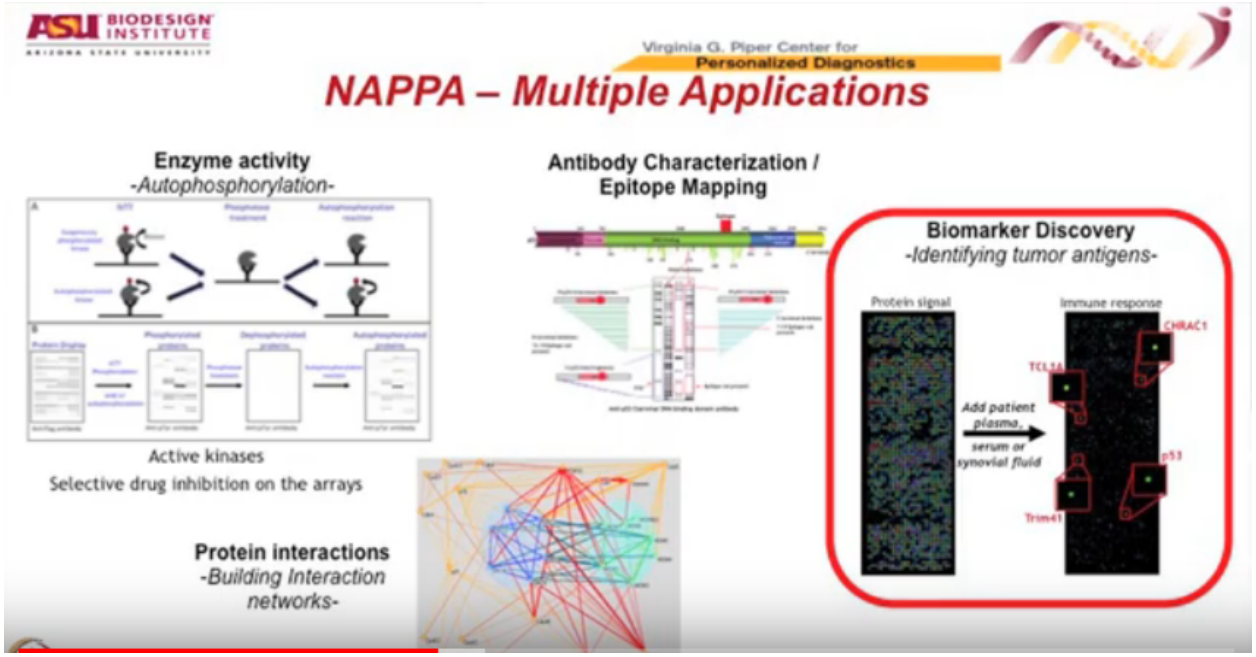
I also showed you this, this is taking a NAPPA, array and we test it for total protein levels, by looking at GST, because all the proteins on the array have a GST tag on their C terminus. So, if we get staining of GST it tells us how much protein is present and if we stain the array for one protein, we get that one protein and I think we've, been through all these sort of advantages of this approach. Okay?

Refer Slide Time :(8: 03)



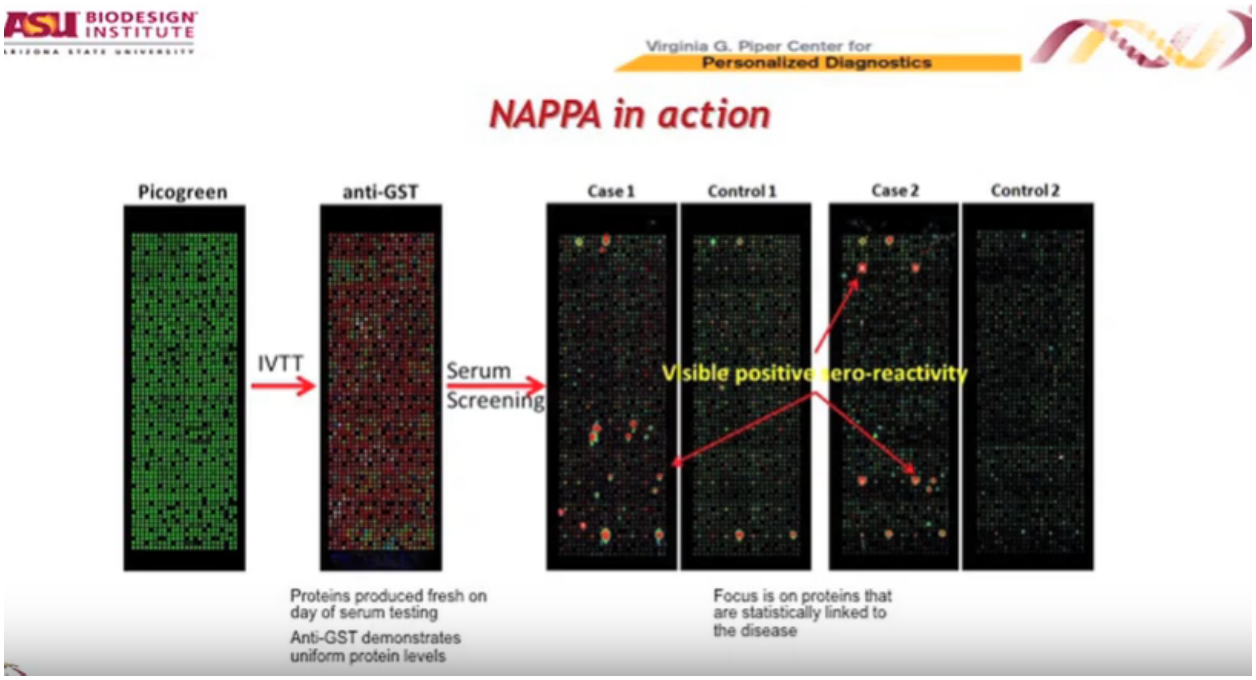
So, let's talk now a little bit about rigor how do we know, that what we're looking at is good well the first thing we want to measure is how, well does the platform express all proteins. Right? So, we're going to take the array, we're going to test it for DNA binding, which tells us how even we printed and we're going to test it for a protein expression, which shows us that we're displaying, all the different proteins and we're displaying them at an, at an reasonable level. And then this is what that looks like? It's summarized quickly but, it shows you four membrane proteins transcription factors kinesis', small large medium and small proteins the, the green line on the bottom is the level of detection. So, if you're above that dotted green line, it means you can detect it that means it's like five 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 it's only a single log. So, no protein is present that's more than ten times than another and the vast majority of proteins are within two fold, of the means. So, they're all very close together. Okay? Right? I should I showed you that so, it's key if you're 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.

Refer Slide Time :(9: 33)



So, these are some of the things that we've done with it about the protein phosphorylation and I'll probably do that in my last talk, I talked a little bit before about mapping protein domains, we talked about the protein interaction studies and now, I'm gonna focus a little bit more on biomarker discovery. All Right? So, this is what we're looking for in biomarker discovery.

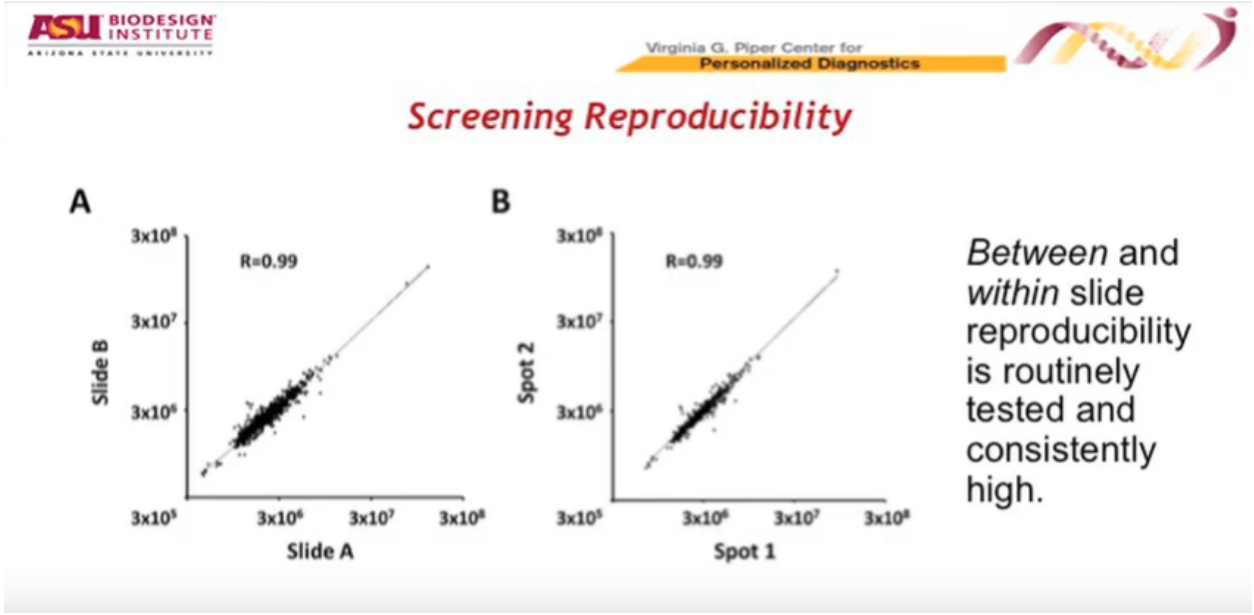
Refer Slide Time :(9: 56)



So, here is the DNA array, here's after making protein and here's 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. Okay? The color by the way is a false color. So, these are not, it's 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 limb, pretty much only a tenfold range by doing different colors, we can cover a larger range. And so, typically blue green is, is blue is weak, green is stronger or orange is stronger, than that well yellow is stronger than, that red is stronger that, orange stronger than, that and red is the strongest. Okay? 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're all very strongly expressed and of course I already mentioned that we've 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's 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, chaperone proteins and a natural I sake 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 serie, 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, 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're, gonna do a clinical study to build, a control sample and run that sample everyday you do the study. So, you can say that, that day everything was working well, if we see that the control sample deviates significantly from what it looked like on the previous days? We'll throw out the data from that day, it's just not worth it you know I'd rather have clean data.
Refer Slide Time :(13: 06)

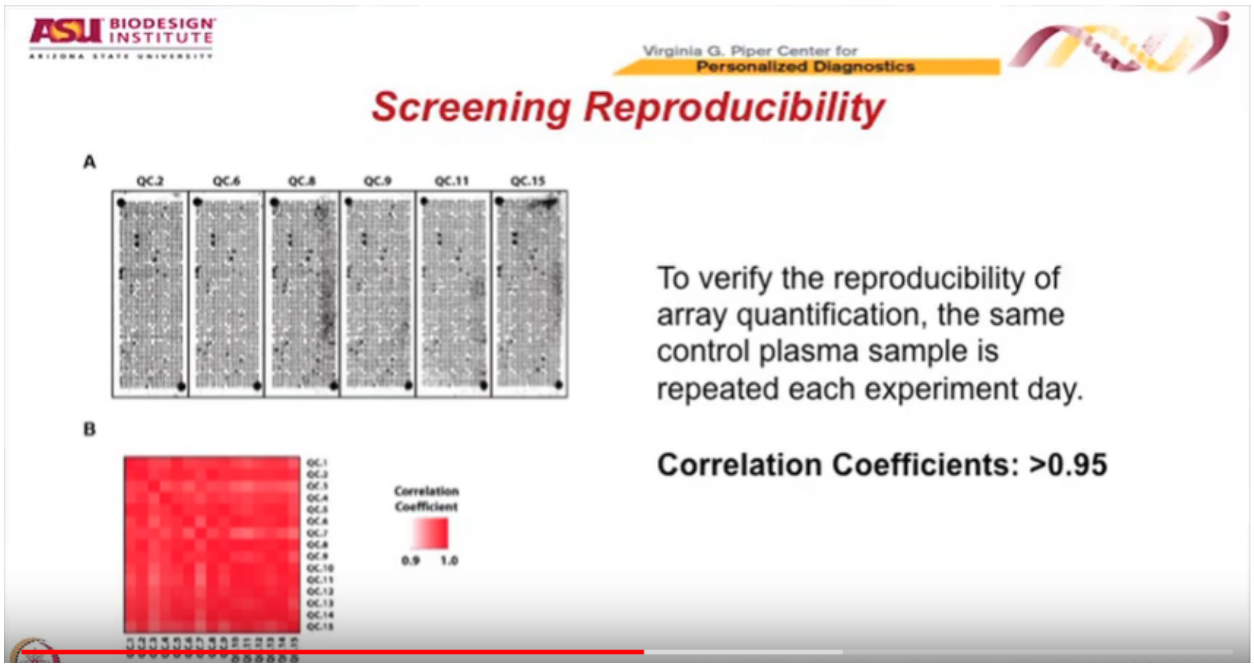
And So, this is what that looks like, here you see comparing day one today two, I think you can see that these two arrays look very similar, in their, in their intensity but, more importantly if you map, all the signal intensities from all these spots, on a dot plot, what you'll see is they all line 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, that's just comparing two days.
Refer Slide Time :(13: 41)



Now, imagine if here's a couple of other examples slide a to slide B here's, 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 similar 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, good correlation from slide to slide.

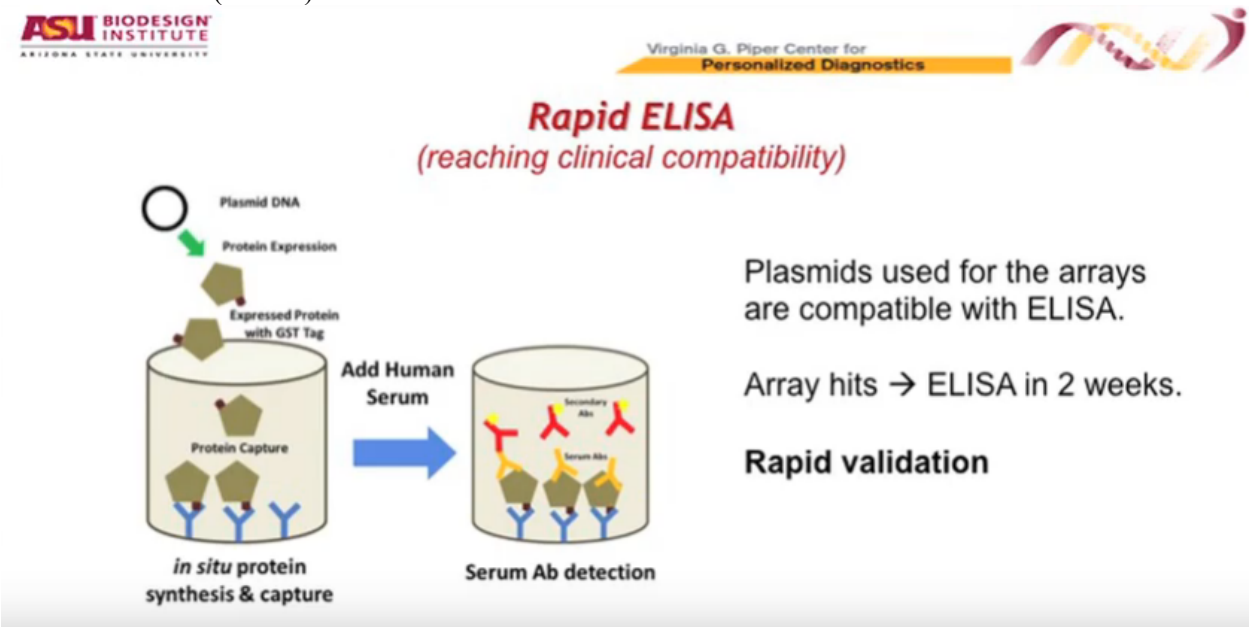
Refer Slide Time :(14: 13)



Okay? Let's look at it for the entire experiment and that's what's 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's what here's 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're bone white, that would be a 90% correlation and if it was solid red, that would be a hundred percent correlation. So, as you can see no matter which two days we compare, nothing gets even close to white almost everything is, 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? It's 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're 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, the observed difference. Right? That we talked about typically you know? In the in the early stages of a biomarker study you're gonna begin by looking at a large, number of possible variables, you're gonna look at as much of the proteome as you can get your hands on. Right? So, that 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's going to correlate with you know? That's going to be predicted for your disease. So, when we begin these, these studies, we're gonna study these days in my lab, maybe fifteen thousand different proteins, that's a pretty good list, now I attached to that list is a price tag. Right? Because by definition, even using NAPPA to get fifteen thousand proteins on a slide, we have to do fifteen thousand DNA mini preps and that's a lot of DNA mini preps it's certainly, easier to do mini preps and it is to make protein but, it's still a lot of mini preps and,

and even if, even if the minute mini preps were two or three American dollars per mini prep, to do that many, many preps is you know? Nearly fifty thousand dollars. Right? Just, just to do that's, that number not to mention the labor and all the time involved in preparing it. So, once we've it, once we've done our initial study where we compared let's say 50 cases, to 50 controls. Right? So, that's a hundred, array sets of 15,000 each, when we get to the next stage, we don't necessarily want to test all 15,000 proteins, we've now eliminated most of those proteins we now know, that of that 15,000, 14,850 of them are probably not good markers, we can toss those out. So, now I don't want to have to use my, my big chip that does all those proteins I want to focus on 150 candidate, markers in my next study. And so, that's where this next platform becomes very helpful now, we can come back to the Eliza assay. Right? And so, the Eliza assay says. Okay? I can make individual proteins at a time, much less costly than making 15,000 protein arrays. Right? But, I can't but, 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 Eliza, without having to do a whole, protein purification from bacteria. Okay?

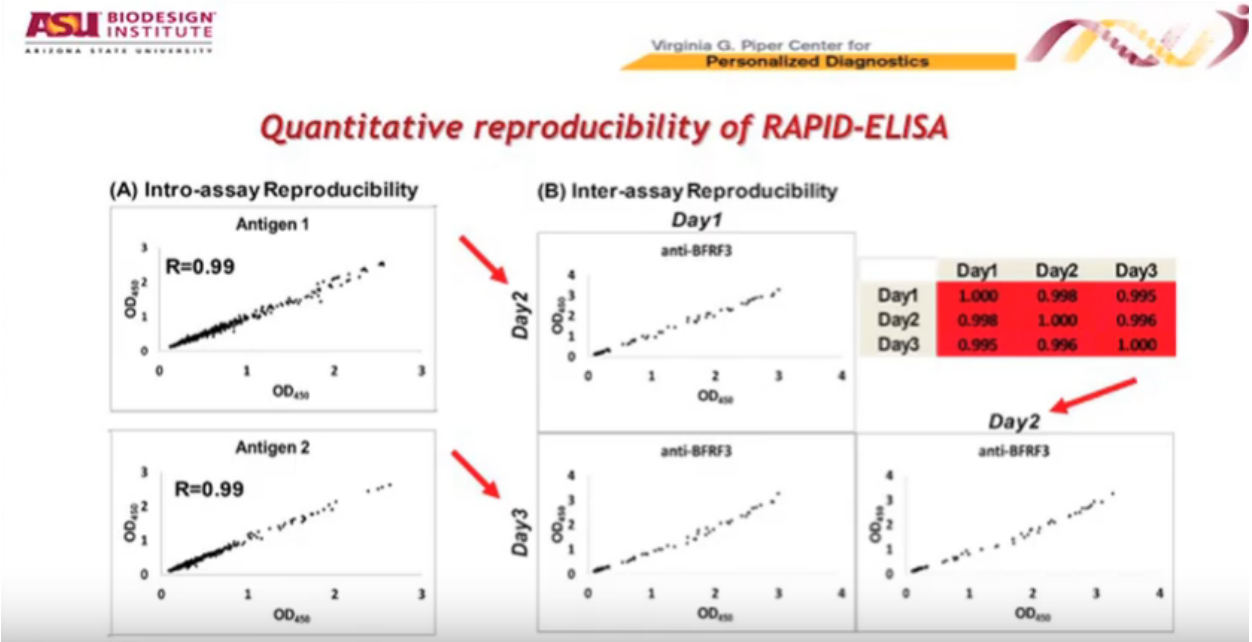
Refer Slide Time :(18: 21)



And we call it, 'Rapid Eliza' and the idea is that we follow kind of a similar, chemistry to the one that we use for NAPA. So, we put the plasmid into the well, we then the, 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're left with a protein displayed in the well. So, if you remember from my second talk where I showed you those, those 96-well plates the sort of early NAPPA that's what we're 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's pretty good especially because we can set up, one of these Eliza's for just about any protein within a couple of weeks one reason we can do that is by definition if, if the protein was detected on a NAPA 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 Eliza. So, the system is immediately compatible with moving

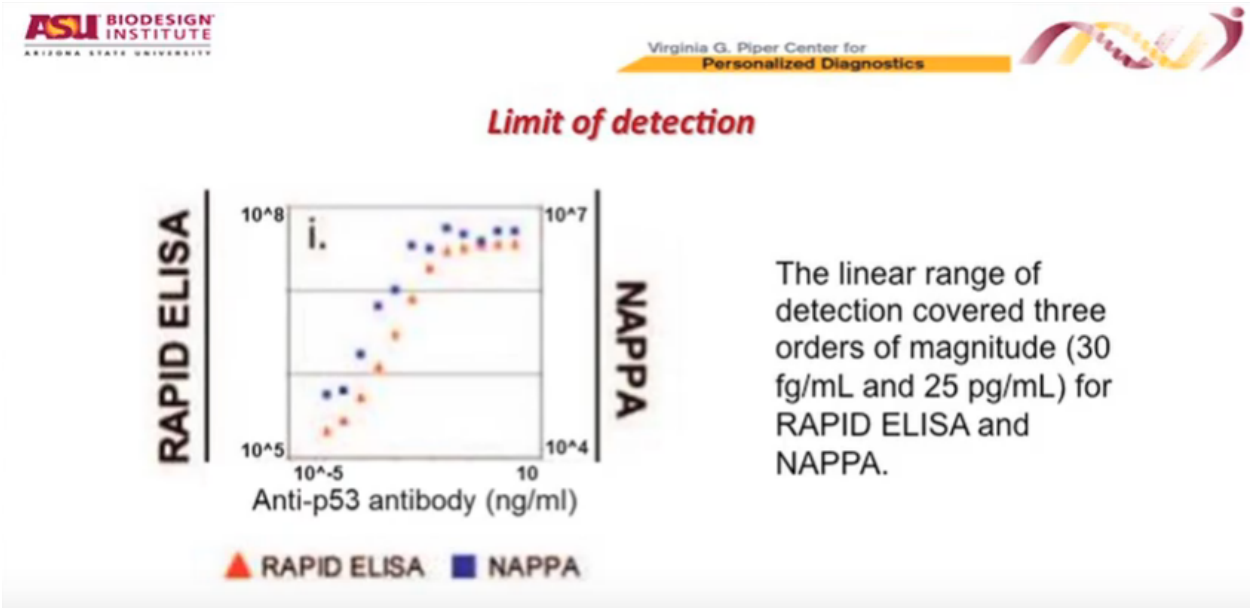
from the array to the Eliza assay. So, we can very quickly set the Eliza up. So we, we add we add the expression plasmid we make the protein we add to the well we capture it, wash away everything we don't need and then we come in with serum to the well and look to see if ,if we get a response.

Refer Slide Time :(19: 57)



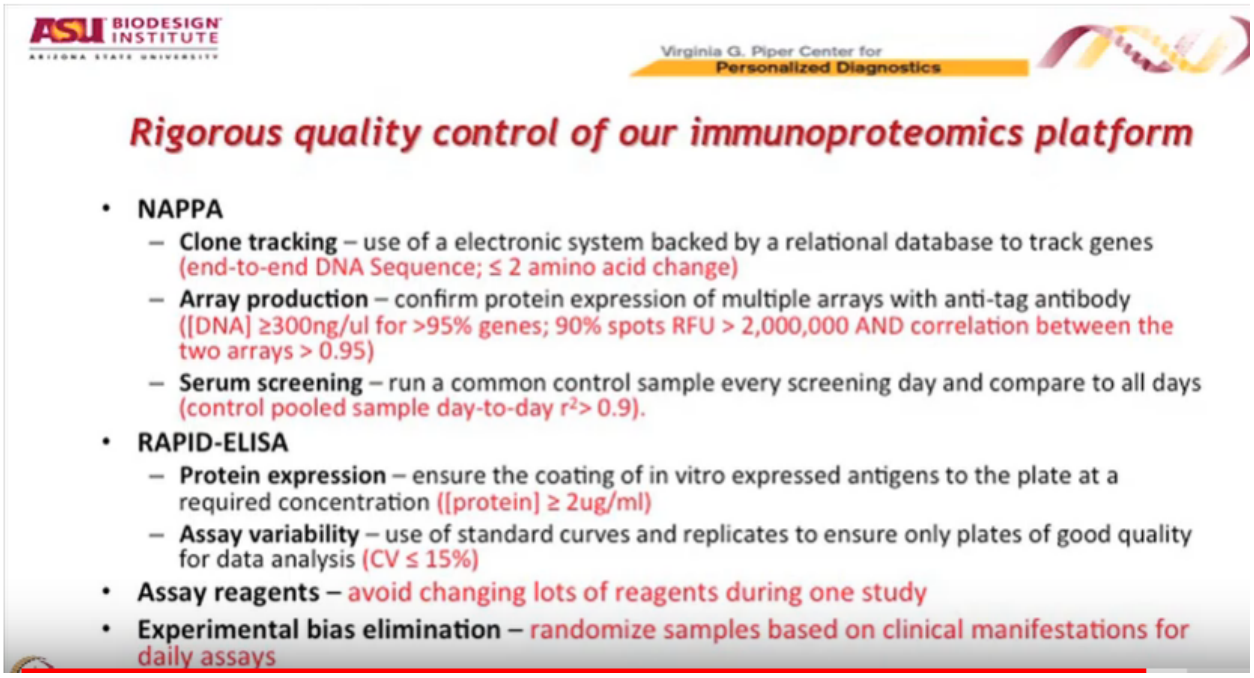
And now, I want to just show you that that the, the Eliza assay is also reproducible and again, you don't want to trust any clinical studies where you can't show reproducibility. So here, here what we're doing is basically looking at a variety of different antigens, these are two different antigens, 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, 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, are typically close to 1.0 ,means that they align very nicely.

Refer Slide Time :(20: 47)



We've also spent a little bit of time looking at our detection limit ,that's what's shown here using a purified amount of anti p53 antibody the, the assays linear from 30 to, 32 grams, to 25 pica grams per mil. So, that gives you some sense this is the range where we can operate very sensitive innocent sensitively. Al Right?

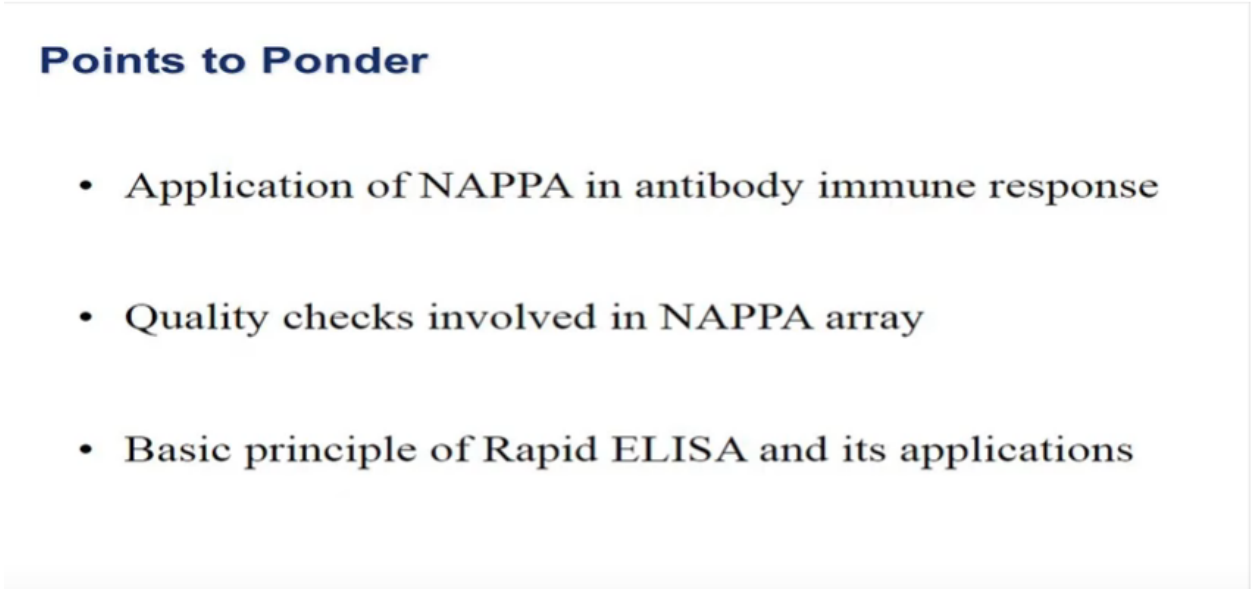
Refer Slide Time :(21: 10)



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, 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 it's not, working that way then you have to, you have to jettison

that step and go back in fix it, before you move on. Right? And so, a good, a good biomarker study involves a lot of QC 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 two amino acids different. So, we will accept up to one 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% of the genes that we're going to print. So, before we even print we look at the whether the plate that has the DNA in it and we make sure that we have an adequate amount of DNA for every, every protein if we don't, 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's protein present and we ensure that 90% 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're checking to make sure that we're adding enough DNA, that we're making enough protein and that the, the features on the arrays from one eight one or eight to another agree with each other. So, we do that for every experiment we do. Okay? And then when we do, serum screening we run the common control sample I mentioned you do that already and we showed that from day to day, the correlation is better than 90 percent, if we're doing a rapid Eliza, 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% 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 do 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 can't tell you how, many students I've had come to the lab and they will do all of the cases today and they do all of the controls tomorrow, they think. All Right? I did it. Right? It's 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 controls 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's 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'll, you'll think you found something really good and then you'll discover a year later after investing all of your time in it, then it was an error because you just didn't, load them in the right order. So, don't, don't miss that I had a postdoc who, who chased his tail for four months, thinking that he had found something really cool and when we got down to it, it was because he ran the case plate first and then he ran the control plate second and the, the control plate sat for ten minutes, while the case plate was getting read and that was the difference. So, you can't do that

Refer Slide Time :(25: 35)



Some sure after attending to the lecture you found this very interesting you have seen how Auto antibody responses for the patient's 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 in a say all this performance has 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 Eliza, a routine technique which is developed to measure the patient responds to any antigen but, at very low cost usually less than a dollar per well. This topic and lecture by Dr. Josh La Baer will be continued in the next lecture. Thank you.