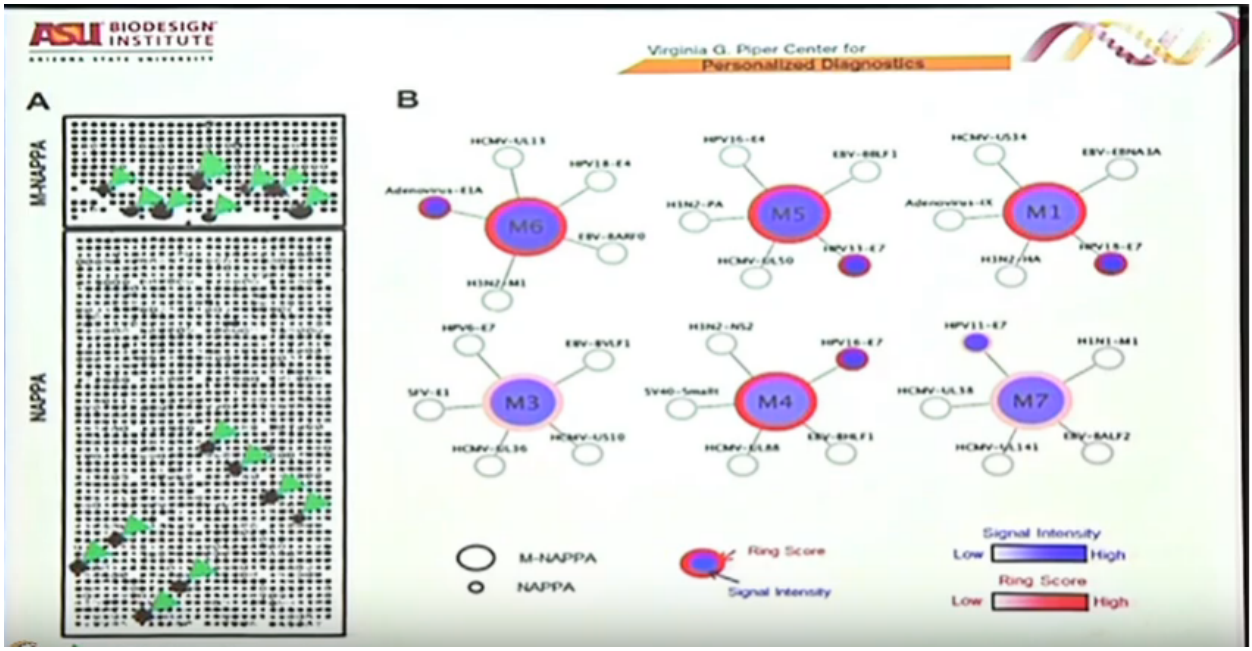
Lecture -10

Using functional proteomics to identify biomarkers and therapeutic targets-I

Welcome to MU codes, on applications of Interactomics, using genomics and proteomics technologies, in last few lectures, you have seen, how the technology platforms like NAPPA? Nucleic acid programmable

protein arrays, could be so useful to perform, high-throughput assays, for the proteins without having, the purified proteins available to you. Just by taking simple CDNA, you can express the protein on the chip and use them, for different type of applications. Today's lecture dr. Joshua labaer is going to continue, discussion about NAPPA technology and mainly ,emphasis will be, on one of the applications about how to use these arrays, for doing research on tuberculosis ,as you know, Mycobacterium tuberculosis. it really affects, large population in whole world and especially more relevant, in Indian context , when we have advent of several, resistant strains of Mycobacterium. Especially, MDR restrains off of TV. So, how to use these array platforms? To do, some sort of novel biomarker based, projects, using NAPPA technology. So, let's welcome, 'Dr. Joshua labaer' to discuss about, applications of Napa, for screening Mycobacterium, tuberculosis .Okay? So, I showed you this, so this is how we analyzed, this is how we analyze the data. So, let me remind you again,

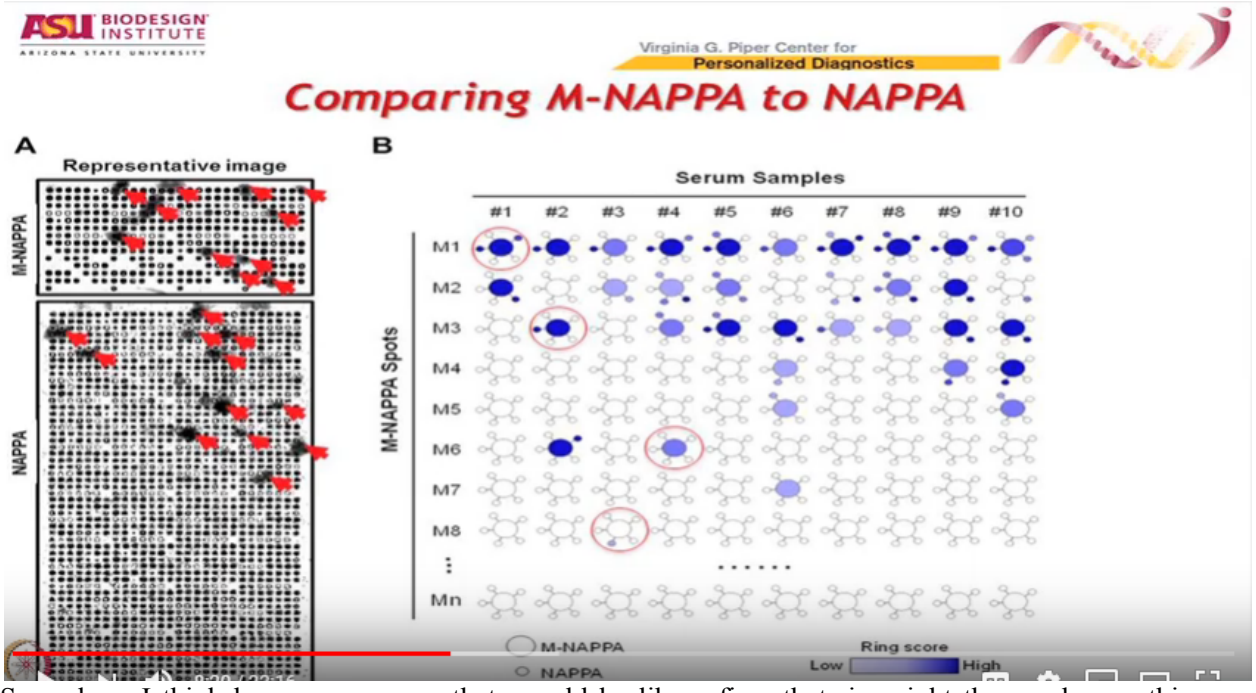
Refer slide time: (2:10)



That this little array up here is multiplexed and this array down here, is all the same proteins, but all as individual spots and so we, we had to, we spent a lot of time, trying to figure out. How do you represent these data? it's always fun when you're doing an experiment the first time and you realize that no one has ever come up with a way to show, what you're looking at ?Because, it's so new. So, what we ended up deciding to do is this kind of ball and stick model, where this ball is representative of the mixed spot and each of these individual circles here, represents the five proteins that are expressed in that spot and then the, the color here, means that, we detected signal at the mix spot and color at any one of these spots, means that, when we tested the individual spot it also gave color. So, if you look at something like this one Right? Here, that suggests that we detected m6, we detected this spot and probably it was the adenovirus a1 a protein that was responsible for that signal, that would be our best guess .Right? here's another ,one and, and probably that protein is responsible for that, here's one that was kind of weak, and notice that we don't see, any of these five spots lighting up .so, that becomes a little bit of a question

mark? So, what would you call that? Thinking back to our statistics from yesterday. It's a possible false positive .Right? The mix, the mix spot says there's a spot, there but when we went to confirm it we couldn't find it. So, that might be a false positive, we don't know, it could be that we just didn't, get good detection here. All Right? Here's another one, also just it's a little bit esoteric here. But, we, we actually look at two qualities of our spots on the arrays, we look at the spot on the array itself, the blue and we also look at the signal intensity of this area around the spot, which we call the ring and a lot of our features have that sort of ring intensity, which is another sign of very strong response.
So, if we see a ring, we usually think it's a very strong positive .Okay? And so, we went through and we, we did some protein interaction studies using this approach ,we also, this array, comprises a variety ,of viral proteins , from common viruses that people are infected with and then we probed it with serum and that for each person would tell us which virus is that person has had before. Right? So, that's shown here and yes, yes, oh. How do we get that value for that? Yeah. That's a little complicated what we do currently? Is that, we have a software application, that pulls up the images and the, the investigator doesn't know, what the spots are? but we have a 5, 5 star scale, for scoring a spot from one, from one to five based, based on this essentially, the size and the look and we have images for each of the five and the investigator looks at the image and then, it's in and then, he scores it or she scores it from a one to five and you do it unbiased, so you don't know, which proteins are, which you just score the, the ring size? it's not, as precise as an instrument doing it ,but it's not bad and so , a very narrow spot, would be like a, one, bigger would be two, three would be a pretty good size, four would be like spilling into the neighboring spots and five would be you know, huge .so, kind of the size and intensity of the spot ,of the ,of the ring. Yeah. AlRight?

Refer slide time: (6:10)

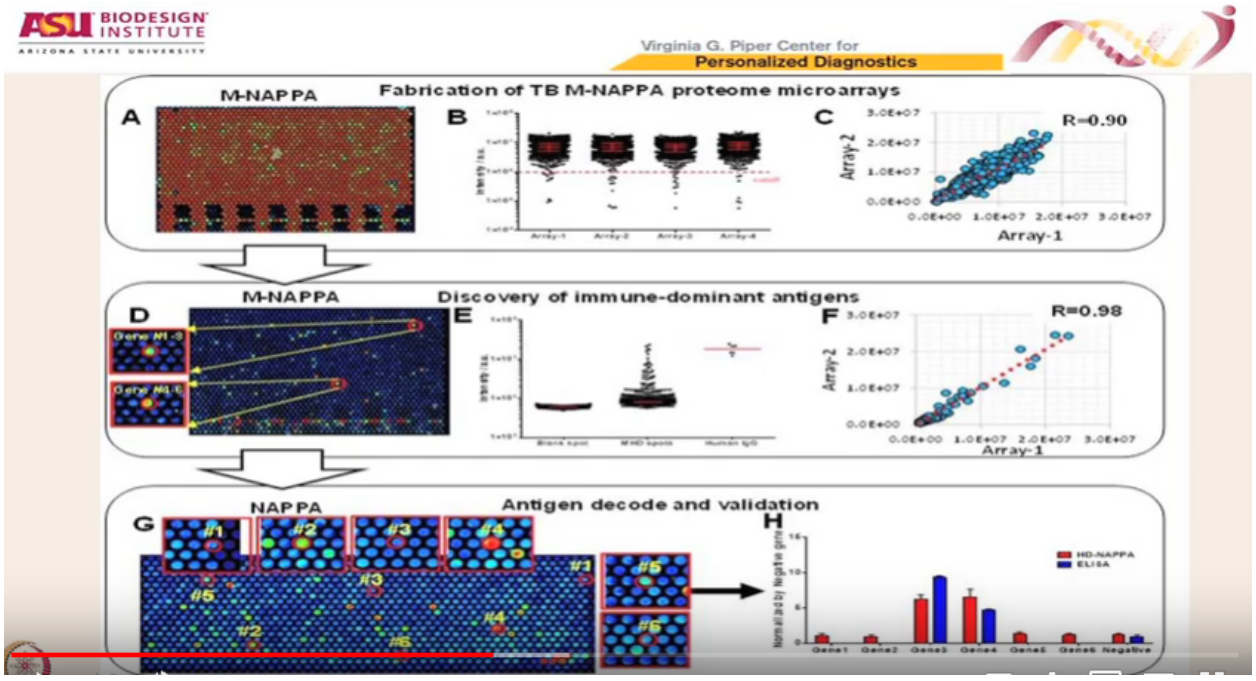


So and so, I think here you can see that, would be like a five, that ring right there, whereas this one probably would be a one or a two. So there you can, see that, the spot there, you can see a little bit of the ring right there, this one is like, that's a five, it's spilling all over ,you know ,this might be a three, up here

that sort of thing. So, it you get used to it after you do a lot of these, so um here ,here you can see some examples ,so each column is a serum sample ,each row is a multiplex spot and then the , the little nodes, around it indicate which proteins were individually ,detected down here. Right? If you look at, this guy here, you it's sort of a standard, what you'd expect? You've got a strong signal in the spot and you also have a spot signal, for that, for one of the proteins in the five. Right? So, presumably this is the one that gave that, signal then here's, another example .where we had a strong spot signal and there were two positives, in that, in that minute. So, that ,that's something that you have to keep in mind is that ,up until now I've been acting like within every mixed spot there's only one target, it's certainly possible that there may be more than one and obviously you hope that the mixed spot ,will give you that signal .Here's, another one.

So, what would, what, what's the concern on this one? It's a false positive. Right? So, this one, that the major spot gave us and it's interesting that the false positives, often are weaker than, the than these guys it's a little bit weaker and then what's that? False negative. Right? Right? So, the individual spot gave us a signal, but the master spot, the mix spot did not. Right? And so these are, the one, this is the one that we worried about the most and we looked at, we looked for these and I think we found, a few percent where that happened, but the vast majority of the time, the mixed spot was sufficient, to find whatever was present in the individual spots. Because, if the two methods agree well, then you feel a lot more comfortable, using the mixed NAPPA, the multiplex NAPPA, over the standard NAPPA. Because, it's so much less expensive.

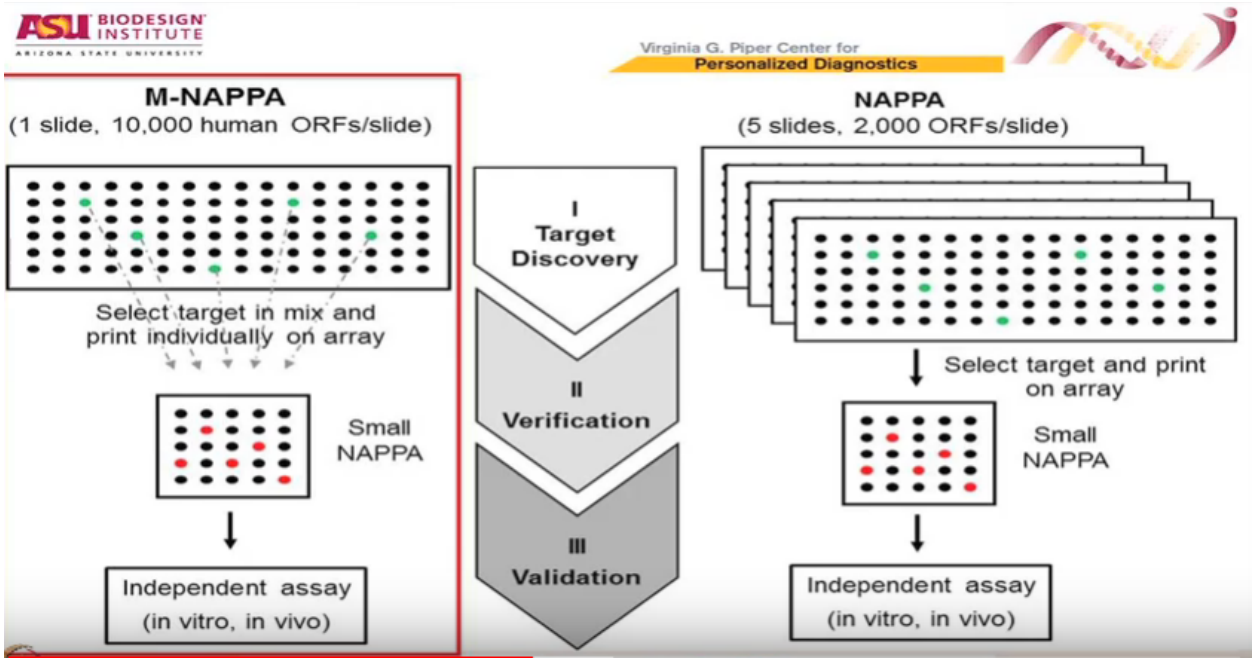
Refer slide time: (8:25)



And so that's, this is kind of the strategy, you, you produce a mixed NAPPA, array up here, you, you get good signal for all these proteins, you show that you have good correlation, you screen for, you screen for antigens that are detected, in the patient let's say and you see a number of spots that light up, you compare

array to, array to make sure that you're getting consistent results ,this is all sort of quality control stuff and then, then after you've got these hits, you take the individual spots that lit up and you break them down, into ,the into individual proteins, the next day to verify that those signals are real and to I in 2d convolute which spots were positive .Okay?

Refer slide time: (9:13)

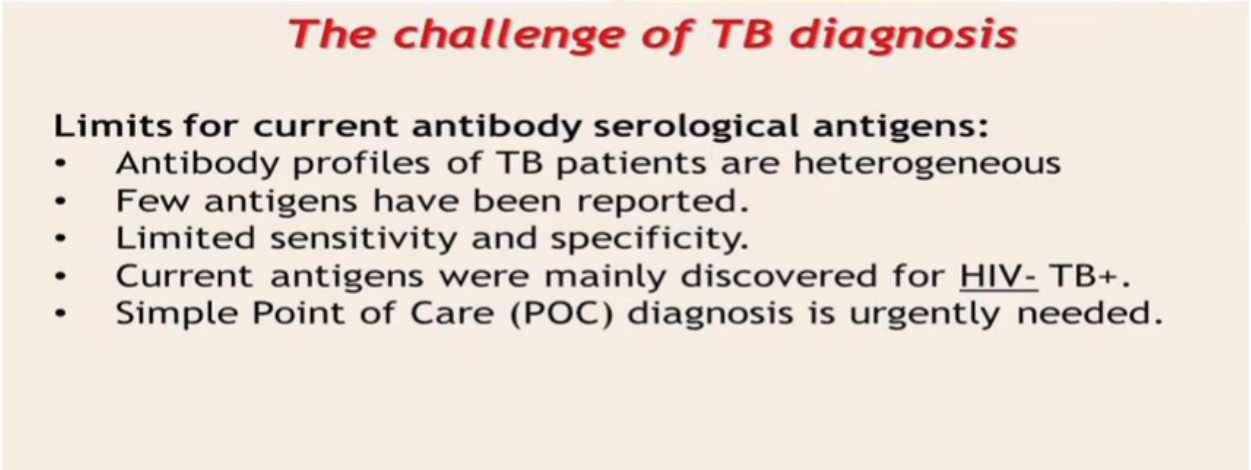


And so, this is kind of the summary of what the mathematics works out. Right? So, in the old way, to do, to do 10,000 proteins, we would have to use five slides and then after we did those five slides, we would have to come back the next day and use a sixth slide, to confirm the hits and, and make sure that they were real. Right? Using the mixed NAPPA, we can do all of the spots on one side and then we just have to come back the next day and do a second slide to confirm that it's a real. So here, we have to do a total of 6; here we have to do only 2. So there's a lot more work on this side, than this side and of course if you're trying to save money, this is definitely a cheaper way to go and yet, you can get, roughly the same numbers. Right?

Refer slide time: (9:58)

Okay? So I don't, I don't have to tell you guys this. Right? People, people this is a big issue ,it's a very huge health issue all over the world, including in this country, one of the big challenges in some parts of the world, is the co-infection of HIV with TB and the diagnostic methods that are available for TB, are already limited in ,in standard TB. But, it's particularly limited in the context of HIV, where the common symptoms and the common molecular studies, don't always apply and so, we were interested in, in studying a little bit whether or not we could identify, good biomarkers, for the detection of active TB, particularly in an HIV -positive population .Okay?

Refer slide time: (10:49)

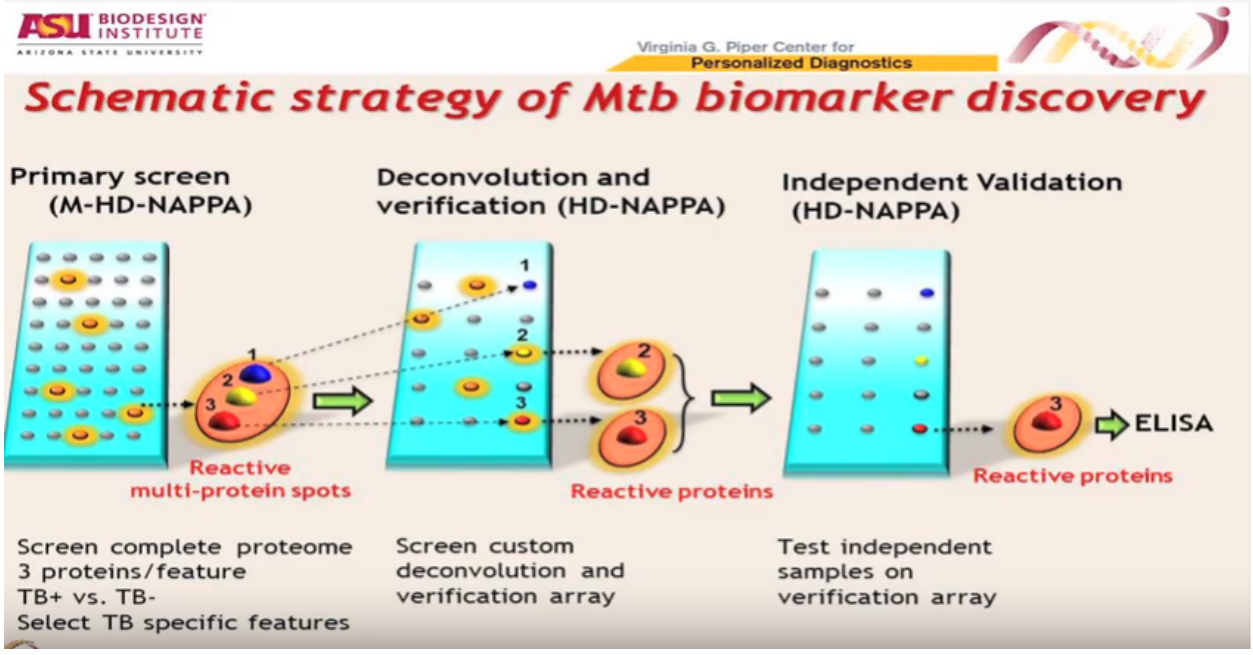


So, these are some of the things that I think you probably already know, those of you, who are aware of a little bit the TV issues or anime profiles from patients could be very different from patient to patient, not

that many antigens have been reported. sensitivity ,specificity are not ideal and even these that have been reported many of them have not confirmed in other studies, we almost always tell you so far I've been in HIV negative individuals and of course ultimately we'd like to get a point of care Diagnostics. so we had done a study, together on a funded grant ,to look at the, the whole TB proteome, screened with serum and all that and for a variety of complicated reasons that particular study, could not be used, there were some issues with mixed, mixed up spots on the array, it's just that you know, normally we don't do this but this was in one case where was a problem and so we reached the end of the study and we couldn't use the data and yet we really wanted to do this study and we want you know, we felt responsible to the agency for publishing it and so we needed to repeat it. But, we were out of money and so we were really operating on a shoestring, we had like no funds at all and we had to figure out a way to study the, the TB —proteome, with no money and so that was where the idea for multiplex NAPPA came up, I was wracking my brain, how can I do this inexpensively?

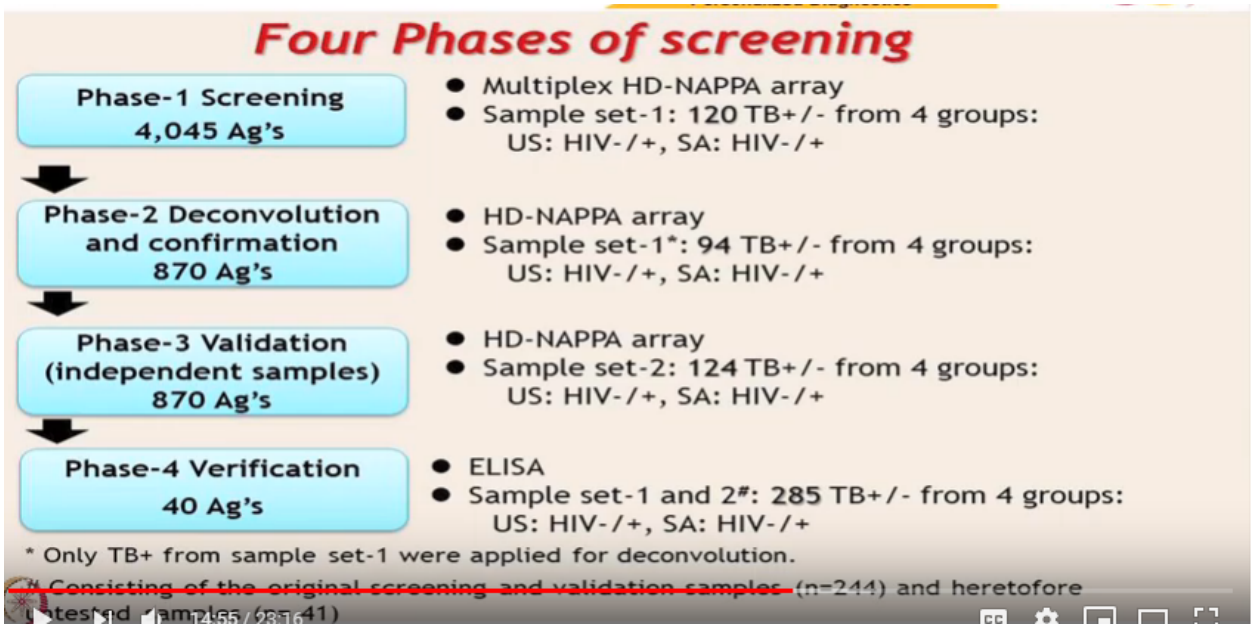
And so it was for that reason that we decided to try the multiplex NAPPA .Okay? because, that way we could get, the entire proteome, of TB, in one quarter of an array ,in fact one array, would be able to do four proteomes at once .Okay? So, that really lowered the cost and of course the other problem with besides the lack of money, was that we were almost out of serum, so we had very little serum to test and of course we needed that to do the study. All Right? So, just as, this is, how we outlined the study? for complicated reasons, we decided not to use, five spots, for five proteins per spot, we used three proteins per spot, part of that was if you remember when I started, I showed you that, that the number of spots that, you can use depends a bit on the, the expected hit rate. Right? And we thought that with TB, particularly in the population from South Africa where we're getting the samples, the hit rate was likely to be higher than just 5%, in part because that population is also co infected by other mycobacterium and those would cross react with the micro bacterial proteins from TB. So assuming that, there was a higher hit rate; mathematically it made more sense to do three spots rather than five.

Refer slide time: (13:37)



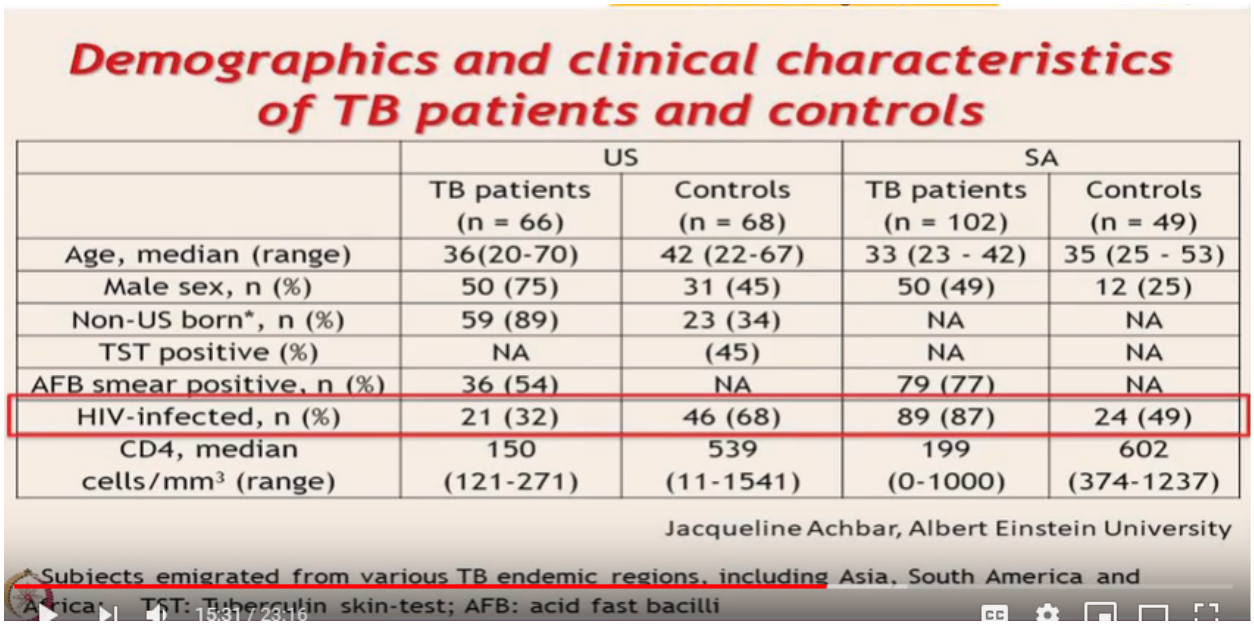
Okay? So we had three, three proteins per spot. Right? And then, we screened that, in with TB sera and positive disease and ,and ,and healthier individual disease, compared that, when we got a hit we took that hit and we divide it into three individual spots, in that secondary verification and deconvolution array, we did this separate array and then from that we got individual proteins that were positive and then we did a third level of validation, by testing different, different samples on a verification array and then finally we took those individual spots to Eliza .so once again, a multi-tiered set of experiments to make sure that whatever hits we observe, really makes sense .Okay?

Refer slide time: (14:28)



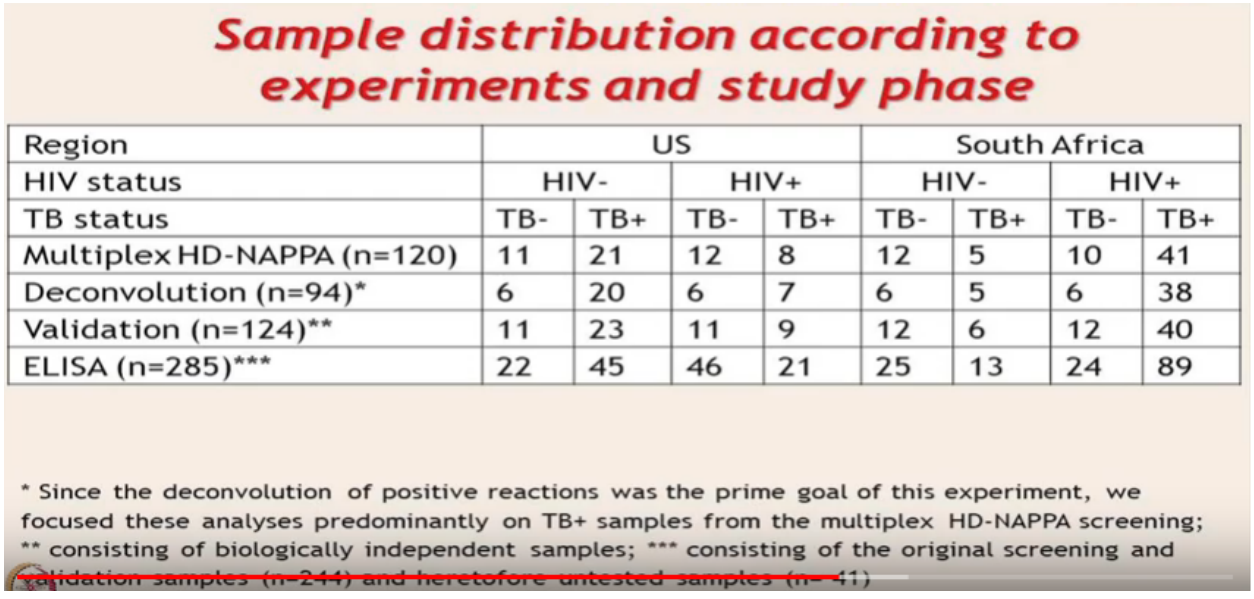
And so this is kind of a summary of that, in phase one we screened, 4000 antigens which is the entire proteome of TB, Phase two we did a deconvolution on what turned out to be about 870 antigens, we did those eight hundreds of the images on an independent set of samples and then we did a Eliza verification, on the best forty of all those alRight? And then this kind of gives you a breakdown, our samples I think, I have a better slide for that part,

Refer slide time: (14:59)



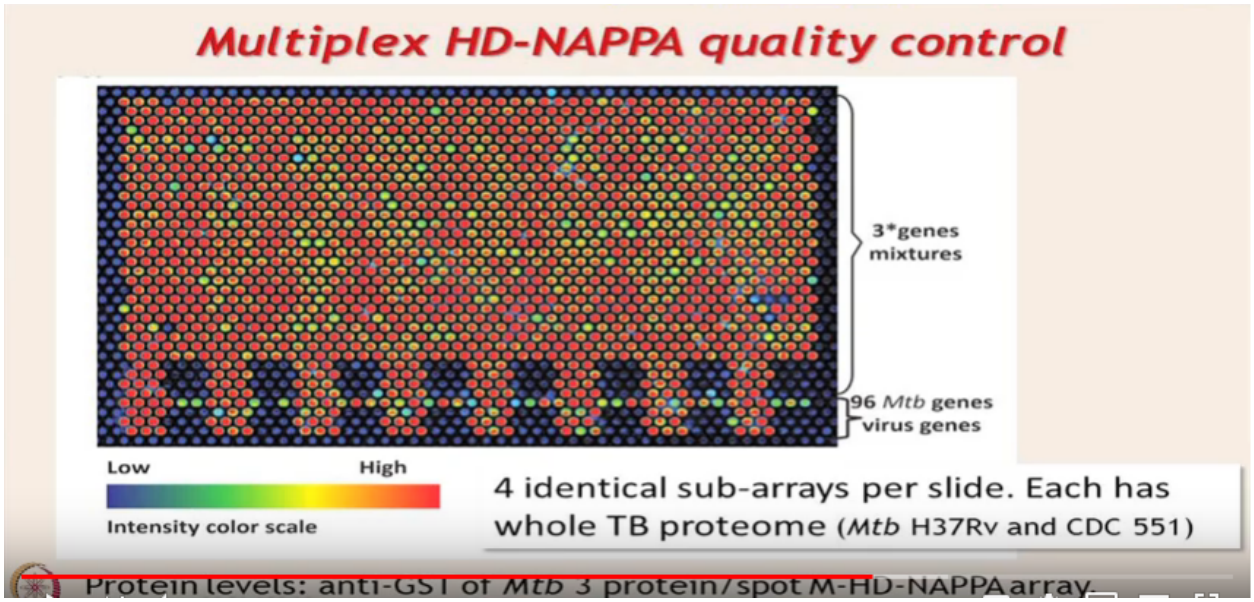
Yeah. This, this is a breakdown of the samples that we had, so we were looking a ,samples that came from the US and Kant samples that came from South Africa, we had T V patients and controls, we also had of these 66 a fraction of them had HIV, of the 68 again fraction of HIV so we had patients who had both who were either HIV positive or HIV negative and did or did not have TB. If that all makes sense ,so it's kind of a complicated clinical design, because we were looking at two factors, TB and HIV and to different countries.

Refer slide time: (15:35)



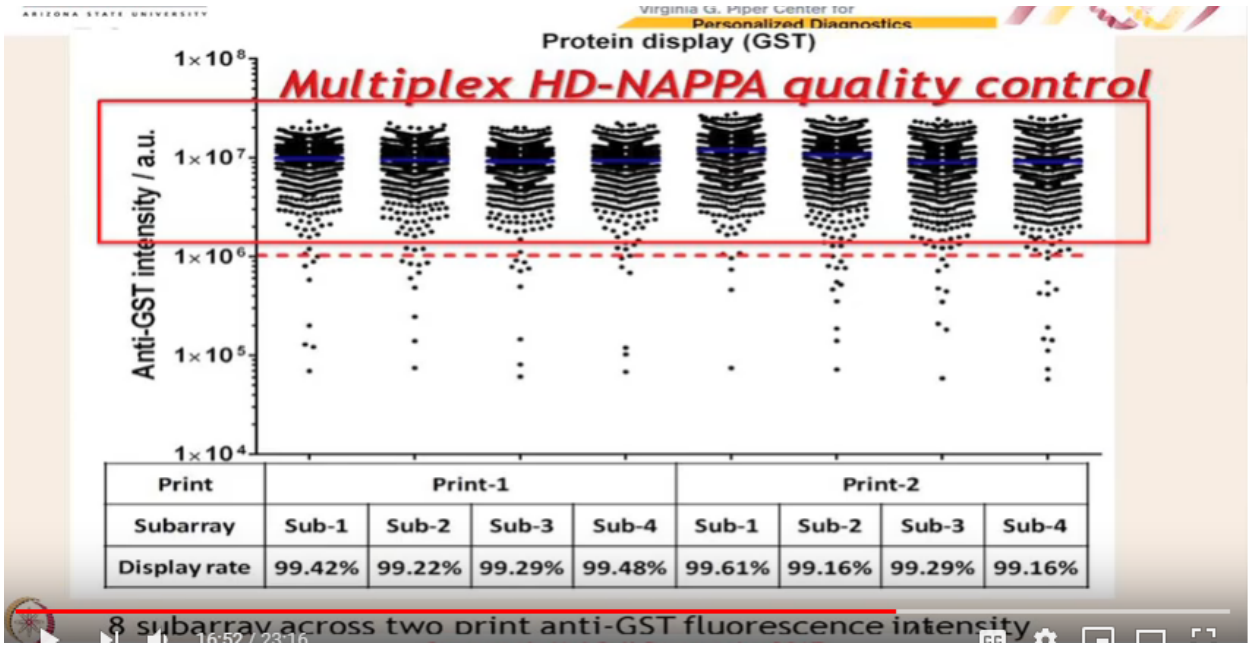
And this is how we broke the samples, into all the different, studies and I this is, published last year in MCP. So I won't, go through all the numbers here. But, you generally get the idea.

Refer slide time: (15:48)



So this is what the array mix look like, when we printed it the total protein. So again we're using, the HD NAPPA and multiplexed. so we've combined the two technologies I told you about, into one experiment today and so these are all the proteins, we also had a subset down here, of individual proteins and viral genes, that did two things for us it gave us a reference spot for individual proteins, it also gave us some, positive controls so that we could make sure that everything was working, because we knew that, most people would have a response, to some of these common viruses .Okay?

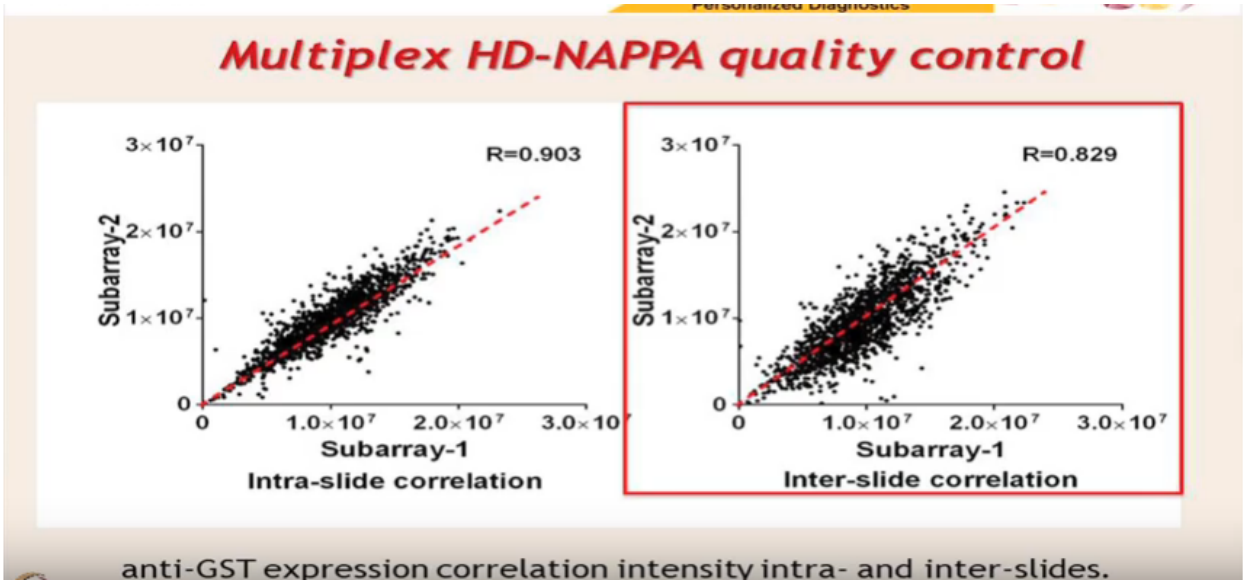
Refer slide time: (16:29)



Print	Print-1				Print-2			
Subarray	Sub-1	Sub-2	Sub-3	Sub-4	Sub-1	Sub-2	Sub-3	Sub-4
Display rate	99.42%	99.22%	99.29%	99.48%	99.61%	99.16%	99.29%	99.16%

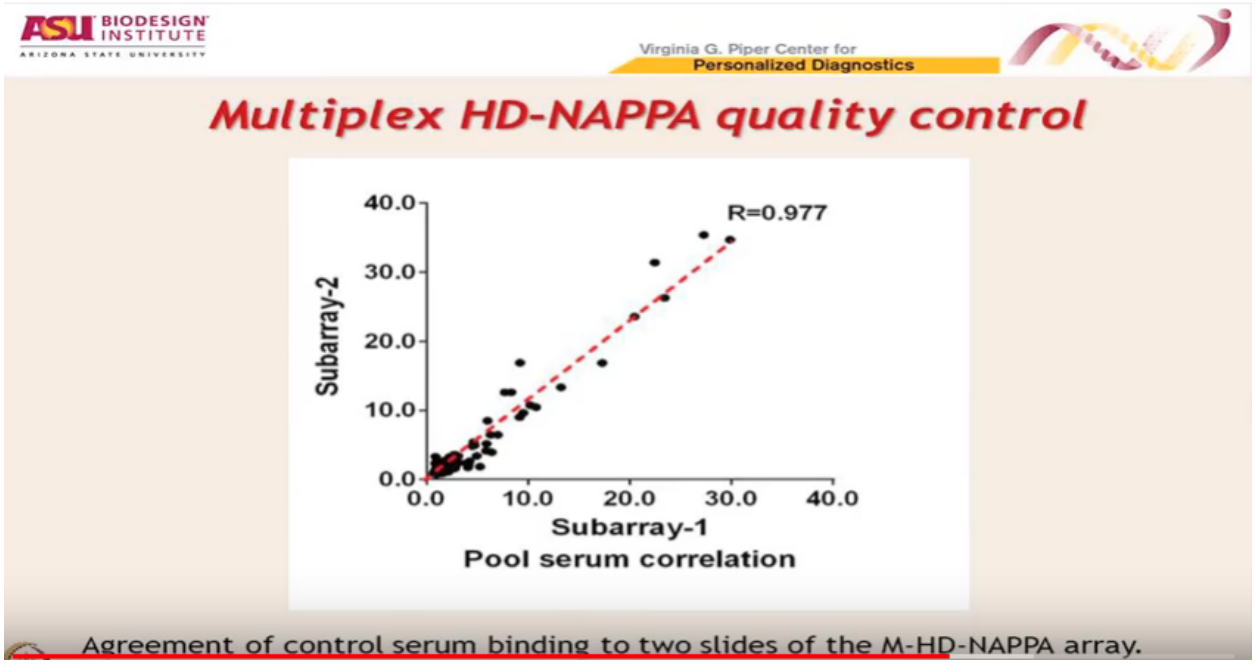
And then this is just to show you, that we got good expression on the arrays, so the dotted line is the cutoff line, everything below that line was considered to be, absent on the array, there was a signal there but it wasn't real, all of this up here, shows that by and large the vast majority of proteins were well expressed on the array and easily detectable and of course that means that we have a good chance at detecting an immune response .Okay? And then this is just once again to show you that .Okay? well I should have mentioned this back here, this what you're looking at Right here, is one quarter of an array, so every array had this set, four times and what that meant is that then we could ,we could screen four patients per array, using special chambers that isolated each of the four chambers, So, so what?

Refer slide time: (17:24)



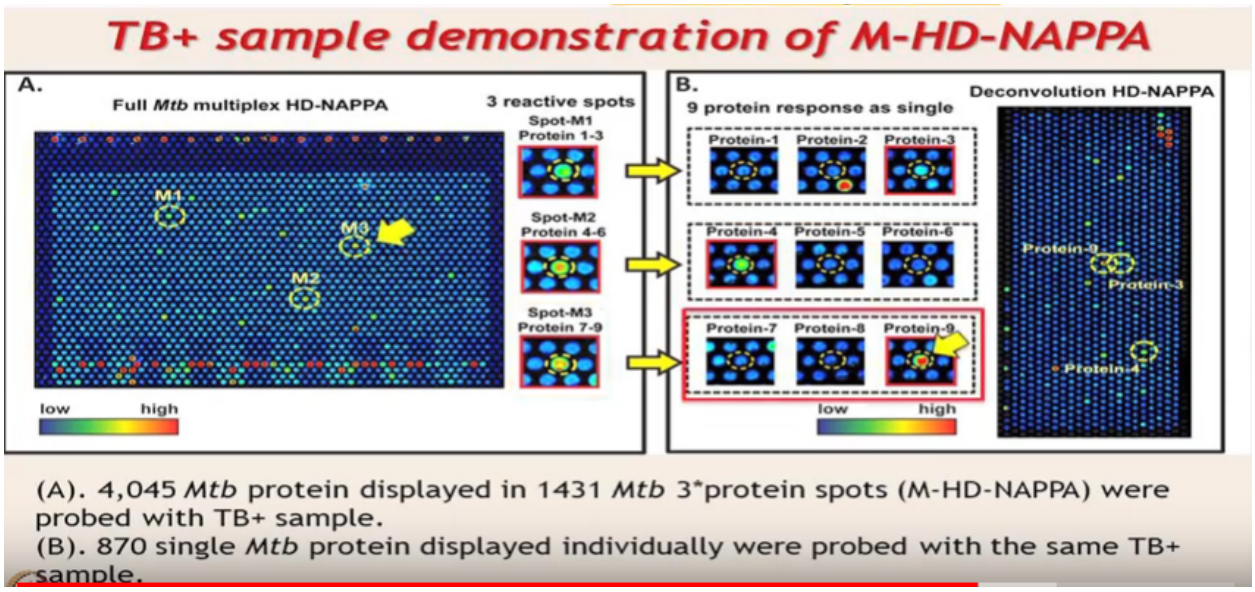
So this is, this what we did here is we compared one of those sub arrays, to another sub array, to make sure that from sub array ,to sub array they were, they reproducible, that's what this is, this is within a slide and then this is looking at two different slides, again to make sure everything is aligned. Right?

Refer slide time: (17:40)



And then when we did our screening experiment ,once again, we want to make sure that we were getting the same result day to day and so here, you see, two different arrays, two different days and again you're getting very good alignment there.

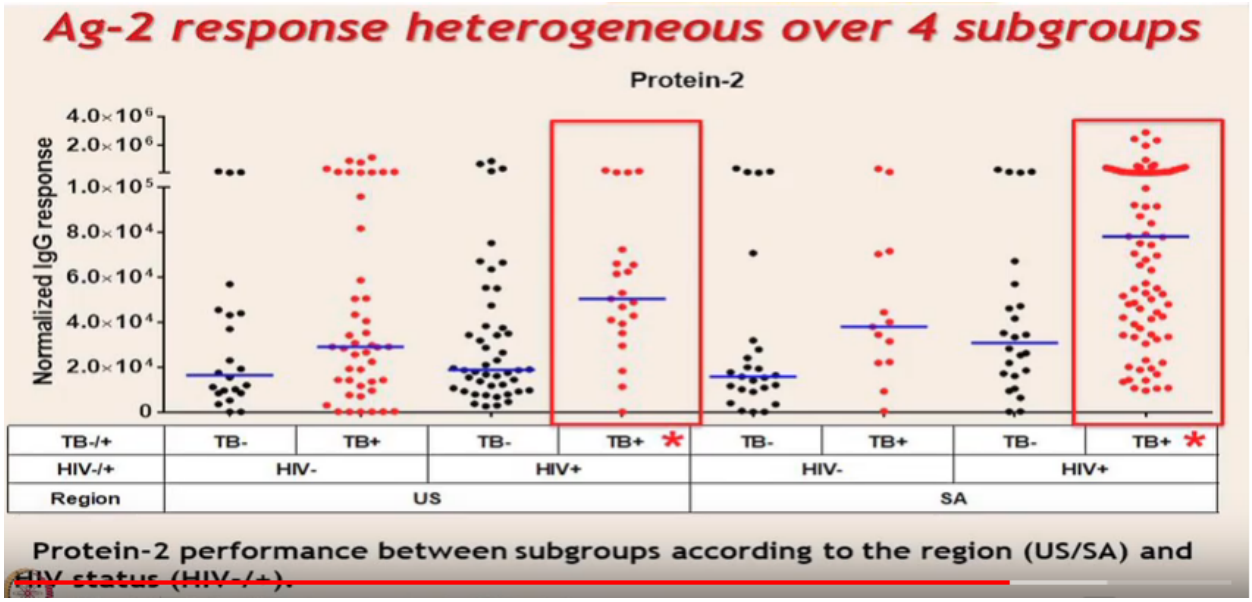
Refer slide time: (17:55)



So this is kind of what we you end up with , here were three spots that they were detected by this particular patient, so here's m1, here's m3,here's m2. Right? then we come over to M one and then on the next day we break it down into individual proteins and what we could see, is that it was protein number three here, that was the protein that lit up for m1, for M 2 it was protein number 4 this protein here that lit

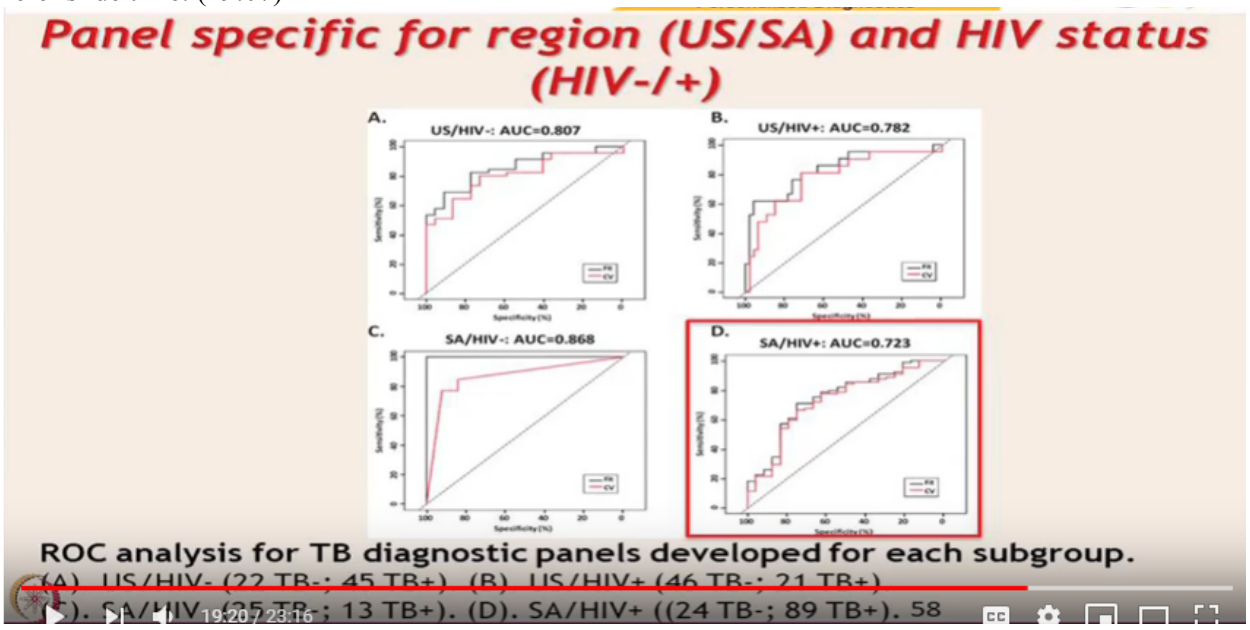
up and for M 3 it was this protein here that lit up. So we're able to deconvolution the next day, yeah. And then, then when all said and done and we analyzed, all the data for all the patients,

Refer slide time: (18:36)



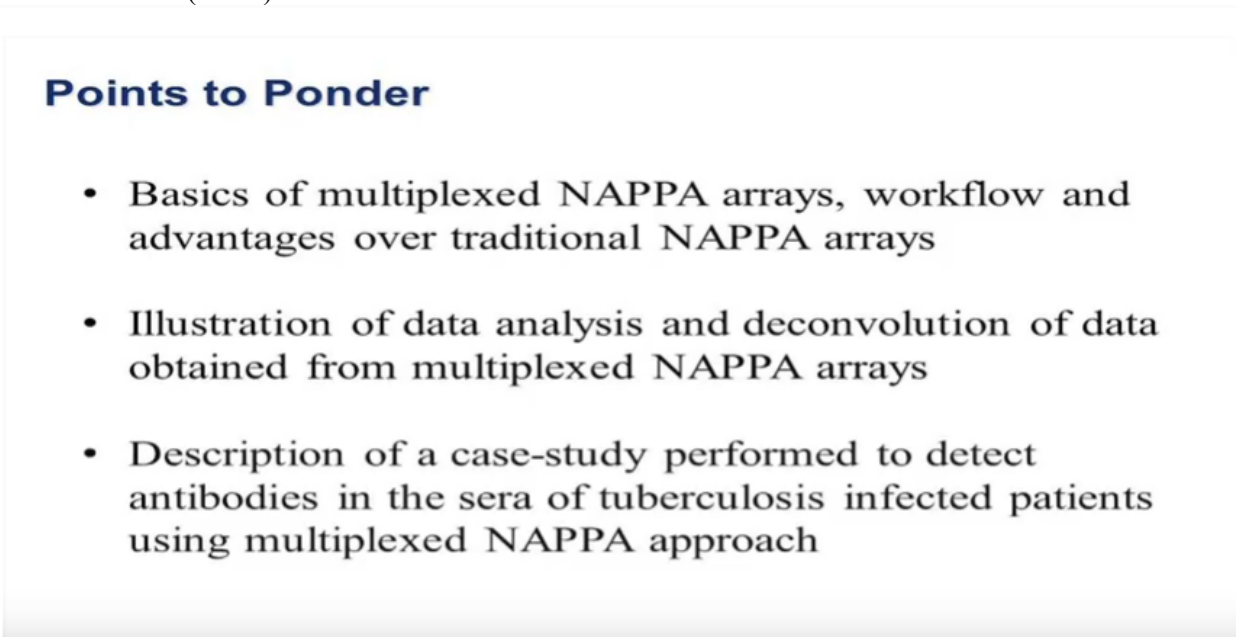
there were some trends that were interesting, here's a protein that's clearly showing a preponderance of signal and TB + over TB negative, in an HIV negative population, that's exactly the kind of marker we were looking for, here's another marker again showing a strong signal in the and even a difference in overall median signal and TB positive, in this case for HIV positive, can sort of see a difference, even in the HIV negative population which is what we were really trying to get to.

Refer slide time: (19:07)



And then by combining those markers together, we were able to get this AUC curve, which is a measure of biomarker quality and you can see it's the curve is not at its strongest and the HIV, in this South African HIV-positive population. But, we're really aiming for this HIV negative population and that looks pretty good. So I think that is what I've got here today, maybe I can stop and take questions. Right? So, if I were to develop a clinical test Right now, we would probably take the top antigens, the overall performing antigens, make more robust assays. Because, these assays are academic lab grade assays they're not you know, commercial grade, make them more robust and then I would probably test that product in each population separately, in order to derive what the cutoff value should be, that's the approach I would take if I were going to commercialize.
It all 40 were tested with Eliza, yeah. And these results are Eliza results, yeah. We're pretty high confident with these, but they're not you know, I mean, if you look at the sensitivity. Right? The sensitivity here is not, outrageous. Right? Because there are you know negatives, TB negatives that have signal now as you know this is challenging. Because, what we call a TB - patient might not be a TB- patient, a lot of these patients have had TB —before and they don't have it right now, or they have other micro bacteria or all kinds of other things could be going on, not as true for the US population as it is for this population, but nonetheless it can still be true and so you do the best you can. But, yeah. So, I would probably include multiple antigens, just because of this issue that there is a little bit of overlap.

Refer slide time: (21:14)



Today's lecture, you've seen, that a new technology platform ,like NAPPA can make research into new areas, so easy and so useful, imagine that expression and purification of all the MTB clones could have been so difficult. but ,now because DR. Labaer group has access to all these clones, for micro bacterium strains, it has very easy for them to prepare the chip, which contains all the genes of Mycobacterium stream and therefore ,no one could use this technology platform, to screen the patient's serum samples who are affected from micro bacterium, of course you have seen ,that they didn't need to look into various type of controls, people who are never having infection, people who are having latent infection, active TB and then also if you could add, you know, the sample population, affected from the various type of

resistant strains, those could provide very novel information and probably new insights for the clinicians to treat these deadly diseases. We'll continue our discussion about, use of NAPPA technology and other protein array platforms, especially for the drug screening in the next lecture. Thank you.