Interactomics Basics and Applications

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

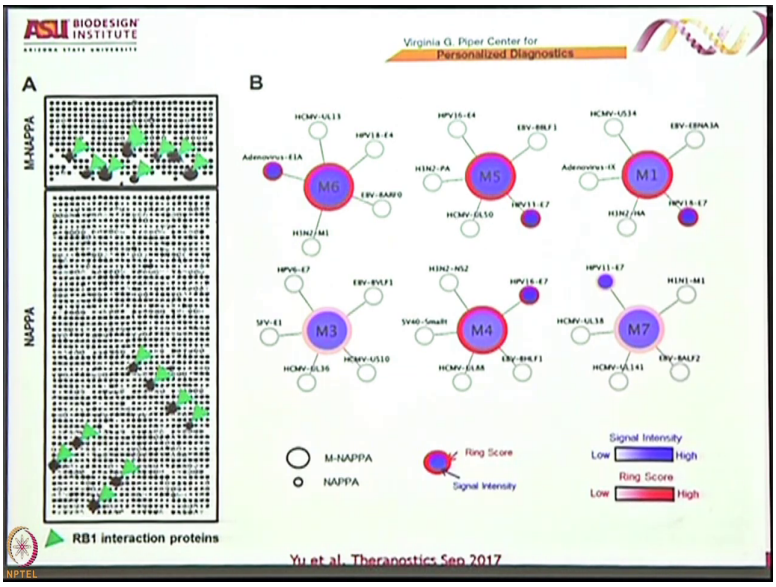
Lecture – 14
Using functional proteomics to identify biomarkers and therapeutic targets-I

In last few lectures, you have seen how a technology platform 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 strains of TB. So, how to use these array platforms to do some sort of novel biomarker based projects using NAPPA technology. So, let us welcome Mr. Joshua Labaer to discuss about applications of NAPPA for screening mycobacterium tuberculosis.

So, I showed you this. So, this is how we analyzed this is how we analyze the data.

(Refer Slide Time: 01:50)



So, let me remind you again that this little array appear is multiplexed and this array down here is all the same proteins, but all as individual spots. And so, we had the we spent a lot of time trying to figure out, how do you represent these data. So, it is fun when you are doing an experiment the first time and you realize that no one has ever come up with a way to show what you are looking at because it is so new.

So, we end 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 5 proteins that are expressed in that spot and then, the color here means that we detected signal at the mixed spot and color at any one of these spots means that when we tested the individual spot, it also get color. So, if you look at something like this one right here that suggests that

we detected M 6, we detected this spot and probably it was the adenovirus E1A protein that was responsible for that signal.

That would be our best guess right. Here is another one and probably that protein is responsible for that, here is one that was kind of weak and notice that we do not see any of these 5 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? So, possible false positive right, the mix spot says there is a spot there, but we went to confirm in we could not find it.

So, that might be a false positive, we do not know we could be that we just did not get good detection here all right. Here is another one also just so little bit esoteric here, but 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, we see a ring, we usually think it is a very strong positive ok.

And so, we went through and 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 with serum and that for each person we tell us which virus is that person has had before right. So, that is shown here and yes.

Student: What is the main basically means the ring?

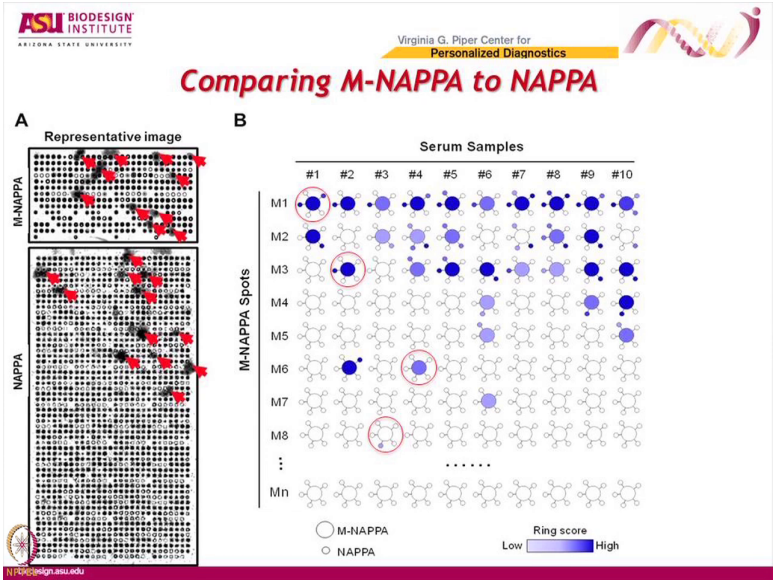
Yes, how do we get the value for that? Yeah, that is a little complicated. What we do currently is that we have a software application that pulls up the images and the investigator does not know what the spots are but we have a 5 stars scale for scoring a spot from 1 from 1 to 5.

Student: Based on intensity?

Based on this essentially the size and look we have images for each of the 5 and the investigator looks at the image and then, it is in and then he scores it or she scores it from a 1 to 5 and you do it on biased. So, you do not know which proteins are which you just score the ring size. It is not as precise as an instrument doing it, but it is not bad. So, a very narrow spot would be like a 1 bigger would be 2; 3 would be a pretty good size 4 would be like spilling into the neighboring spots and 5 would be you know huge.

So, kind of the size and intensity of the spot of the ring. Yeah, all right.

(Refer Slide Time: 05:50)



So, it is so I think here you can see that would be like a 5 that ring right there; whereas, this one probably would be a 1 or a 2. 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 is a 5; its spilling all over you know this might be a 3 up here that sort of thing. So, it you get used to it after you do a lot of these.

So, here you can see some examples. So, each column is a serum sample; each row is a multiplex spot and then, the little nodes around it indicate which proteins were individually detected down here right. If you look at this guy here you, it is sort of a standard what you would expect you got a strong signal in the spot and you also have a spot signal for that for one of the proteins in the 5 right. So, presumably this is the 1 that gave that signal, then here is another example, where we had a strong spot signal and there were 2 positives in that minute. So, that that is something that you have to keep in mind is that up until now, I have been acting like with in every mix spot. There is only one target. It is certainly possible but there may be more than one and obviously, you hope that the mix spot will give you that signal. Here is another one. So, what would what is the concern on this one?

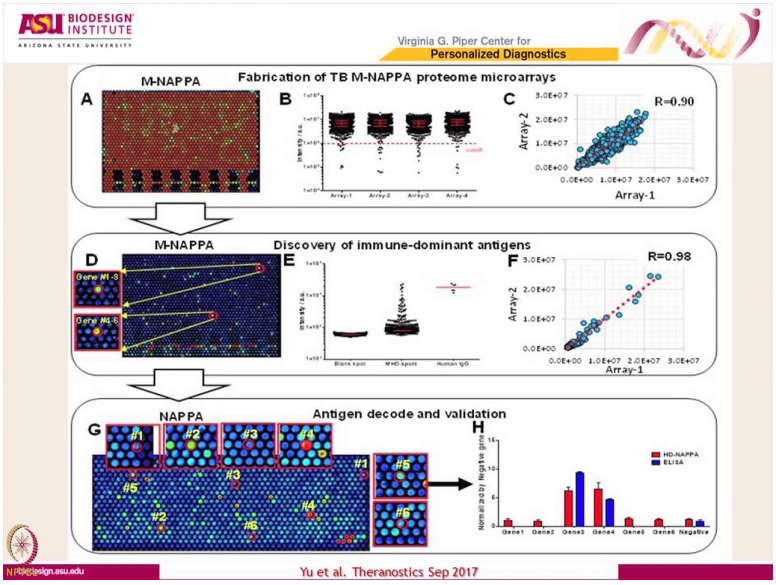
Student: It is a false.

It is a false positive right. So, this one that the major spot gave us and it is interesting that the false positives often are weaker than the than these guys it is a little bit weaker and then, what is that?

Student: False negative.

False negative, right? So, the individual spot gave us a signal, but the master spot the mix spot did not right and so, these are the 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 mix spot was sufficient to find whatever was present in the individual spots because if the two methods agree well, then you feel lot more comfortable using the mix NAPPA, the multiplex NAPPA over the standard NAPPA because it is so much less expensive.

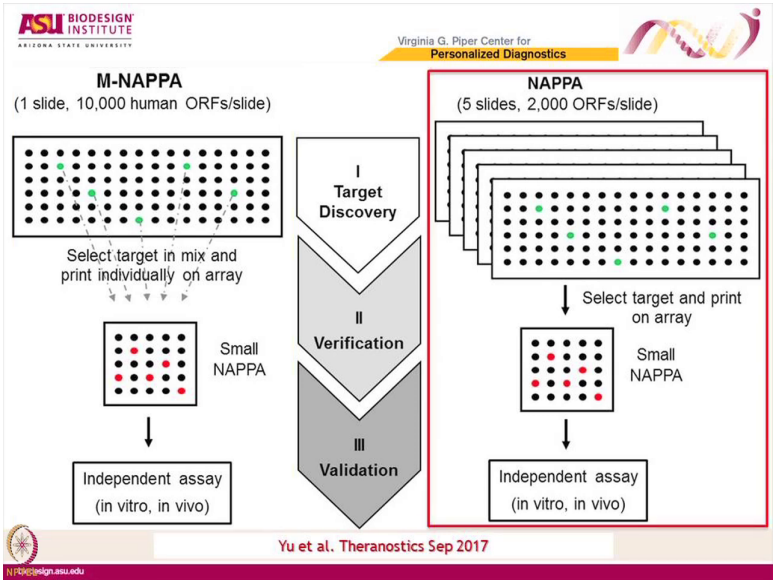
(Refer Slide Time: 08:03)



And so, that is this is kind of the strategy, you produce a mixed NAPPA array up here, you get good signal for all these proteins you show that you have good correlation, you screen for you screen for antigens row detected in the patient let us say and you see a number of spots that light up, you compare array to array to make sure that you are getting consistent results.

This is all sort of quality control stuff and then, after you have 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 into deconvolute which spots were positive ok.

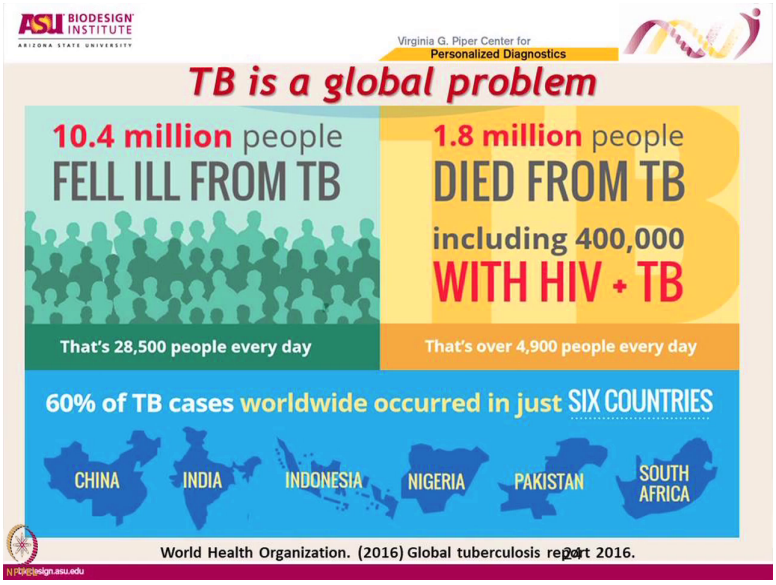
(Refer Slide Time: 08:52)



And so, this is kind of the summary of what the mathematics works out right. So, in the old way to do 10000 proteins, we would have to use 5 slides and then, after we did those 5 slides, we would have to come back the next day and use a sixth slide to confirm the hits and make sure that they were real right. Using the mixed NAPPA, we can do all of the spots on one slide, and then we just have to come back the next day and do a second slide to confirm that it is a real.

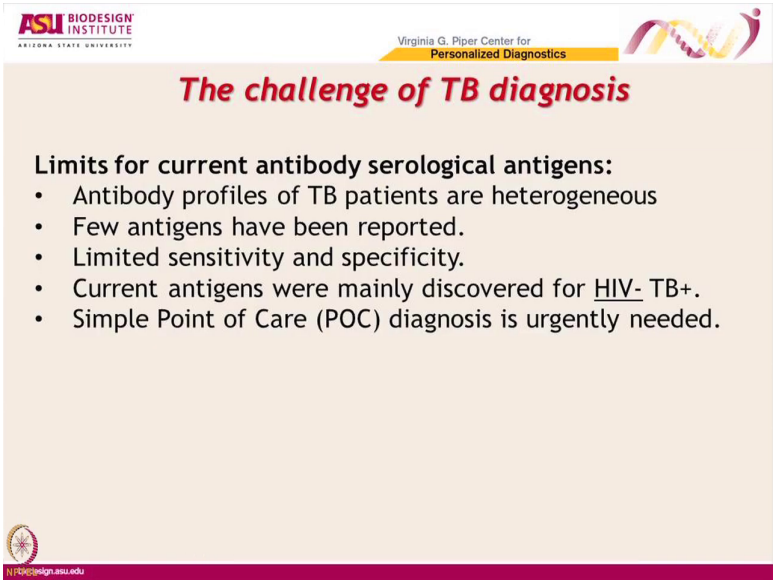
So, here we have to do a total of 6; here we have to do only 2. So, there is a lot more work on this side than this side and of course, if you are trying to save money, this is definitely a cheaper way to go and yet you can get roughly the same numbers right ok.

(Refer Slide Time: 09:38)



So, I do not have to tell you guys this right. People, this is a big issue; it is 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 standard TB, but it is a particularly limited in the context of HIV, where the common symptoms and the common molecular studies do not always apply. And so, we were interested 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 ok.

(Refer Slide Time: 10:28)



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 TB issues, but antibody profiles from patients can 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 say so far have been in HIV negative individuals and of course, ultimately we would like to get a point care diagnostics.

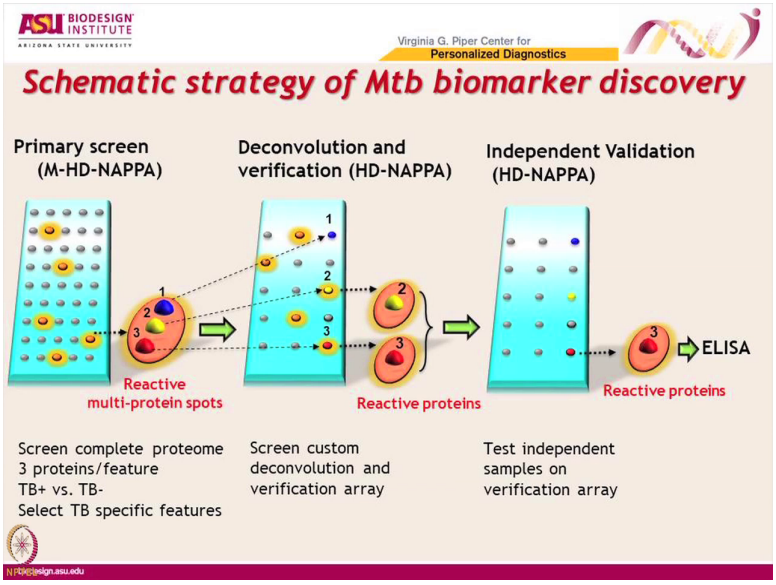
So, we had done a study together on a funded grant to look at the whole TB proteome screened with serum and all that and for a variety of complicated reasons that particular study could not be used. There were some issues with mixed up spots on the array. It just that you know normally we do not do this, but this was in one case where it was a problem and so, we reached the end of the study and we could not 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 need to repeat it, but we are 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 TB proteome with no money and so that was where the idea from multiplex NAPPA came up. I was racking my brain, how can I do this inexpensively.

And so, for that reason that we decided to try the multiplex NAPPA 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 4 proteomes at once ok. 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 5 spots for 5 proteins per spot; we used 3 proteins per spot.

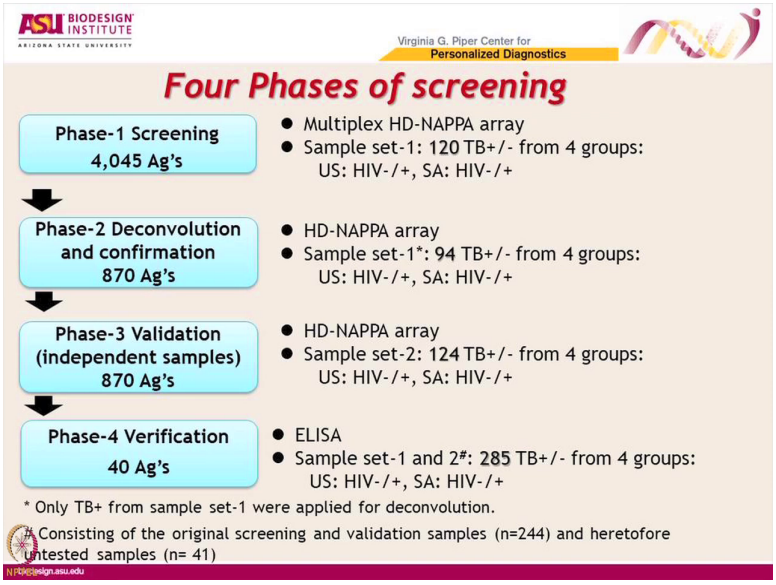
Part of that was if you remember when I started I showed you that the number of spots that you can use depends a bit on the expected hit rate right. And we thought that with TB, particularly in the population from South Africa, where we are getting the samples the hit rate was likely to be higher than just 5 percent in part because that population is also co infected by other micro bacteria 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 3 spots rather than 5 ok.

(Refer Slide Time: 13:17)



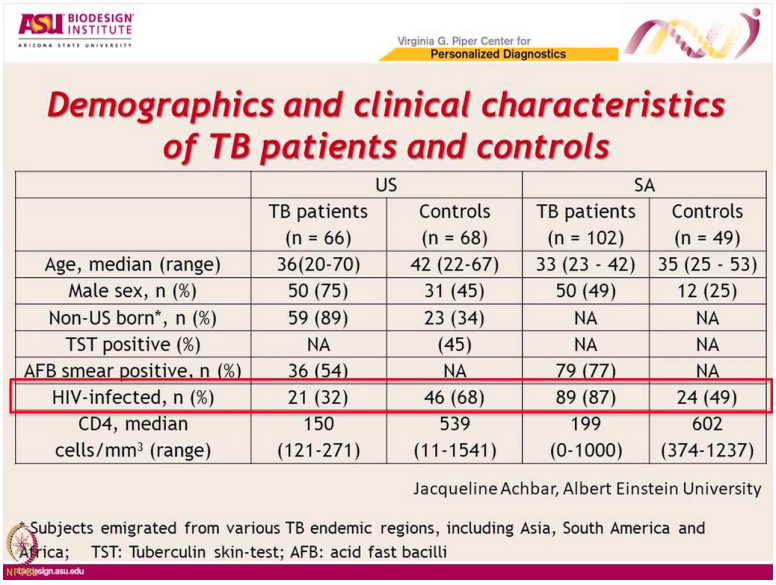
So, we had 3 proteins per spot right and then, we screened that in with TB sera and positive disease and healthier individual disease compared that. When we got a hit we took that hit and we divided into 3 individual spots in that secondary verification and deconvolution array. We did this separate array and then, from that we got a individual proteins that were positive and then we did a third level of validation by a testing different samples on a verification array and then, finally, we took those individual spots to ELISA. So, once again a multi-tiered set of experiments to make sure that whatever hits we observe really make sense ok.

(Refer Slide Time: 14:08)



And so, this is kind of a summary of that. In phase 1, we screened 4000 antigens which is the entire proteome of TB. Phase 2, we did a deconvolution on what turned out to be about 870 antigens. We did those 870 antigens on an independent set of samples and then, we did a ELISA verification on the best 40 of all those all right and then, this kind of gives you a breakdown our samples I think I have a better slide for that part. Yeah, this is a breakdown of the samples that we had.

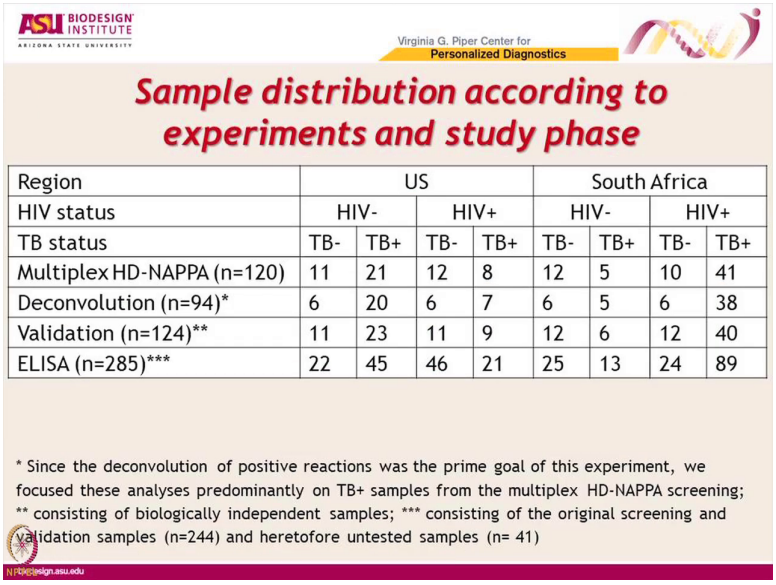
(Refer Slide Time: 14:39)



	US		SA	
	TB patients (n = 66)	Controls (n = 68)	TB patients (n = 102)	Controls (n = 49)
Age, median (range)	36(20-70)	42 (22-67)	33 (23 - 42)	35 (25 - 53)
Male sex, n (%)	50 (75)	31 (45)	50 (49)	12 (25)
Non-US born*, n (%)	59 (89)	23 (34)	NA	NA
TST positive (%)	NA	(45)	NA	NA
AFB smear positive, n (%)	36 (54)	NA	79 (77)	NA
HIV-infected, n (%)	21 (32)	46 (68)	89 (87)	24 (49)
CD4, median cells/mm^3 (range)	150 (121-271)	539 (11-1541)	199 (0-1000)	602 (374-1237)

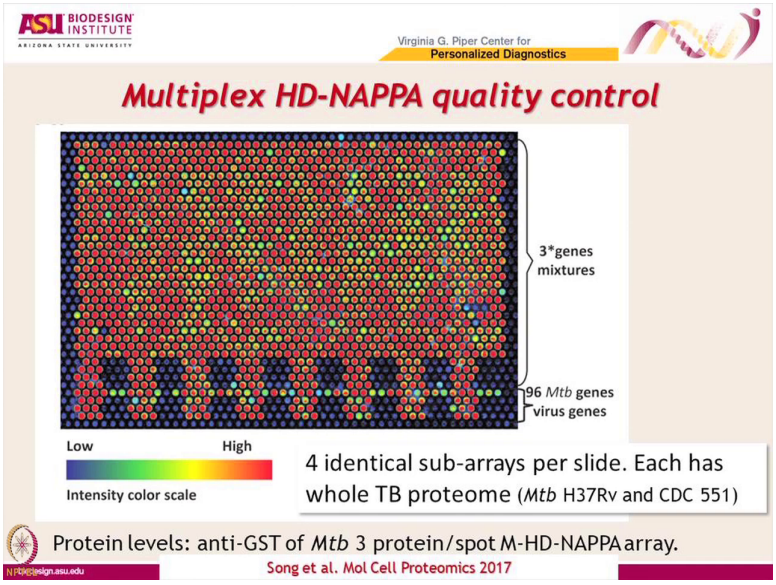
So, we were looking at samples that came from the US and samples that came from South Africa. We had TB patients and controls we also had all of these 66 a fraction of them had HIV; of the 68 again a fraction about HIV. So, we had patients who had both who either HIV positive or HIV negative and did or did not have TB if that all makes sense. So, it is kind of a complicated clinical design because we were looking at two factors; TB and HIV and two different countries.

(Refer Slide Time: 15:14)



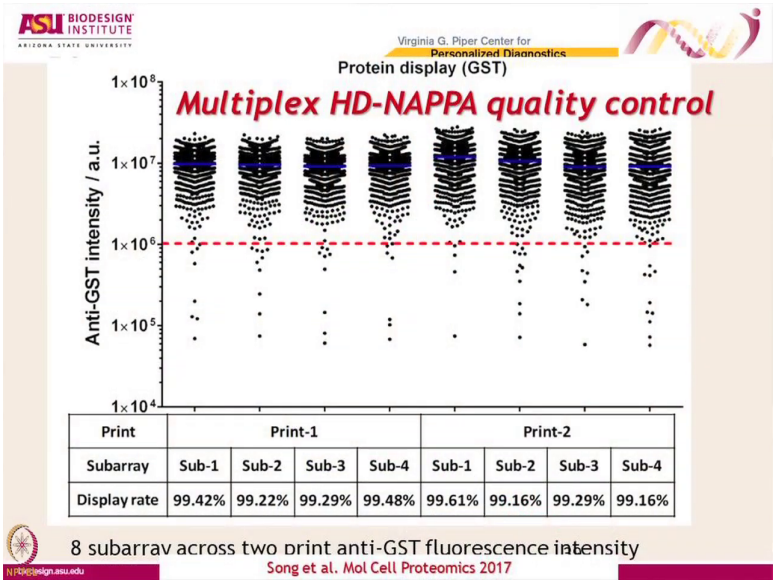
And this is how we broke the samples into all the different studies and I this is published last year in MCP. So, I would not go through all the numbers here.

(Refer Slide Time: 15:27)



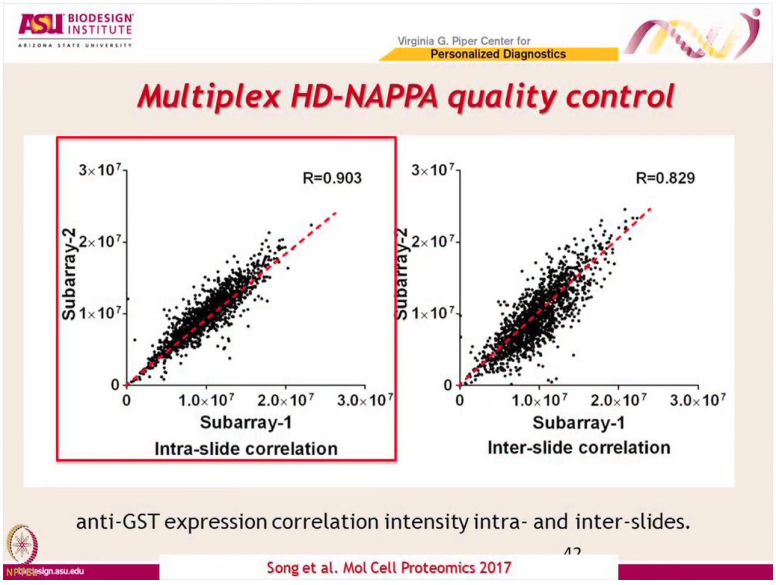
But you generally get the idea. So, this is what the array mix look like when we printed it; the total protein. So, again we are using the HD NAPPA and multiplexed. So, if combined the two technologies; I told you about into one experiment today and so, these are all the proteins. We also had a subset 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 response to some of these common viruses ok.

(Refer Slide Time: 16:09)



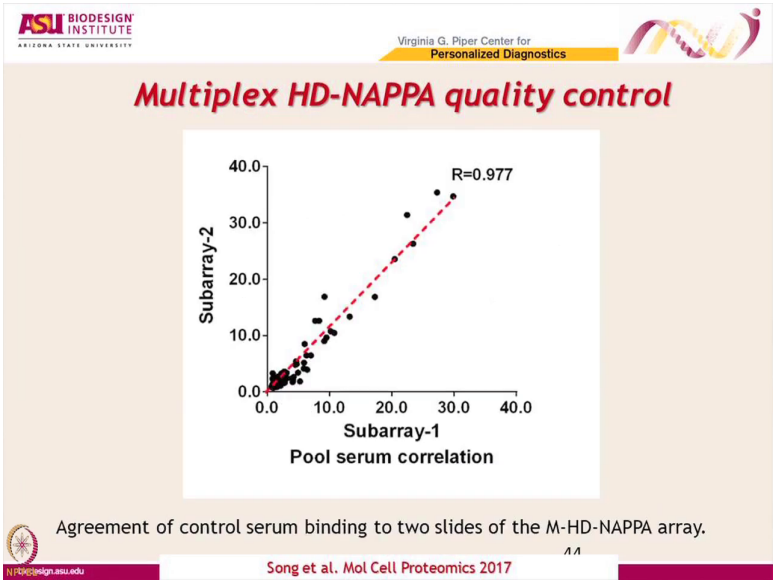
And then, this is just to show you that we got good expression on the arrays. So, the dotted line is the cutoff line everything below that line was considered to be absent on the array. There was a signal there, but it was not real. All of this up here shows that by and large the vast majority of proteins 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 ok.

(Refer Slide Time: 16:36)



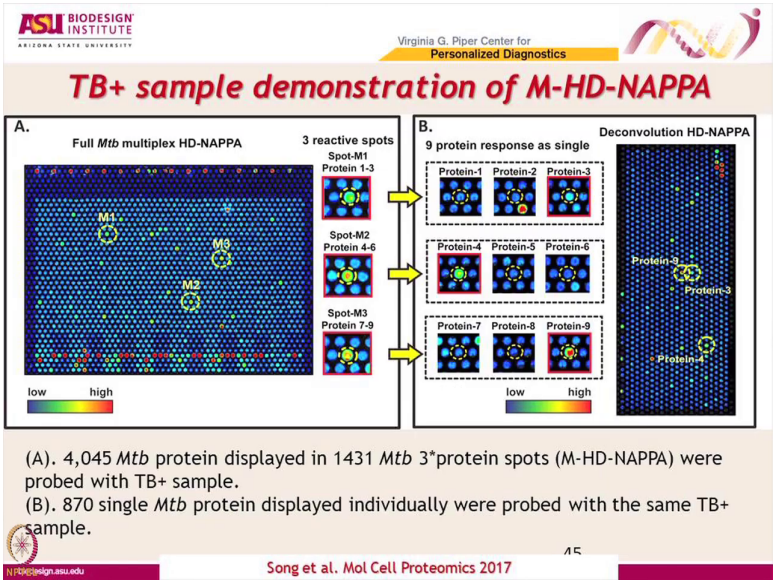
And then, this is just once again to show you that I should have mentioned this back here. This what you are looking at right here is one quarter of an array. So, every array had this set 4 times and what that meant is that then we could we could screen 4 patients per array using special chambers that isolated each other 4 chambers. So, once, 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 reproducible that is what 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: 17:19)



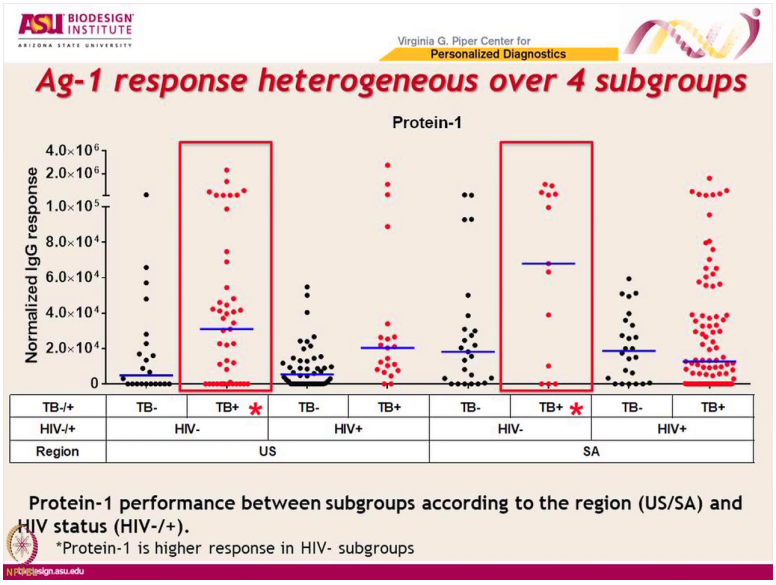
And then, when we did our screening experiment, once again, we wanted to make sure that we were getting the same result day to day and so, here you see two different arrays, two different days and again you are getting very good alignment there.

(Refer Slide Time: 17:34)



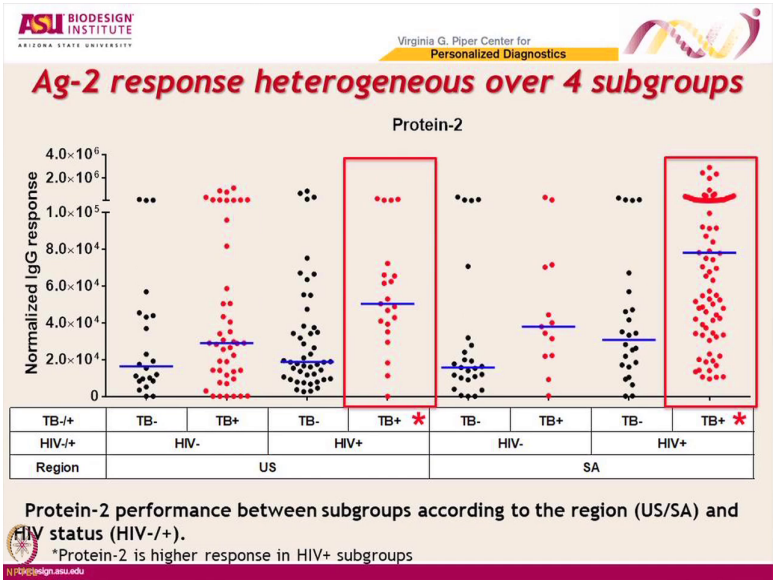
So, this is kind of what we you end up with; here were three spots that they were detected by this particular patient. So, here is M1; here is M3; here is M2 right. Then, we come over to M1 and then, on the next day we break it down into individual proteins and what we could see is that it was protein number 3 here, that was the protein that lit up for M1. For M2, it was protein number 4, this protein here that lit up. And for M3, it was this protein here that lit up. So, we are able to deconvolute the results in the next day. Yeah.

(Refer Slide Time: 18:00)



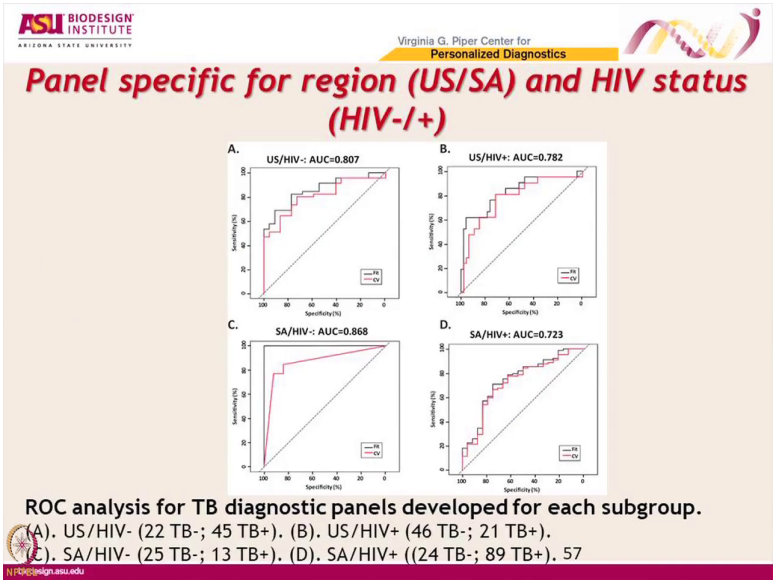
And then, when all said and done and we analyzed all the data for all the patients, there were some trends that were interesting. Here is a protein that is clearly showing a preponderance of signal and TB positive over TB negative in an HIV negative population, so it is exactly the kind of marker we were looking for.

(Refer Slide Time: 18:29)



Here is another marker again showing a strong signal in 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: 18:45)



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 its the curve is not at its strongest and the HIV in this South African HIV positive population. But we were really aiming for this HIV negative population and that looks pretty good. So, I think that is what I have got here today maybe I can stop and take questions.

Student: Which can be used worldwide; so, how would be actually narrow down that what would be in your mixture of a (Refer Time: 19:26)?

Right. So, if I would a 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 are 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 is the approach I would take if I were going to commercialize it.

Student: But sir out of 40 number also tested with the ELISA or?

All 40 were tested with ELISA.

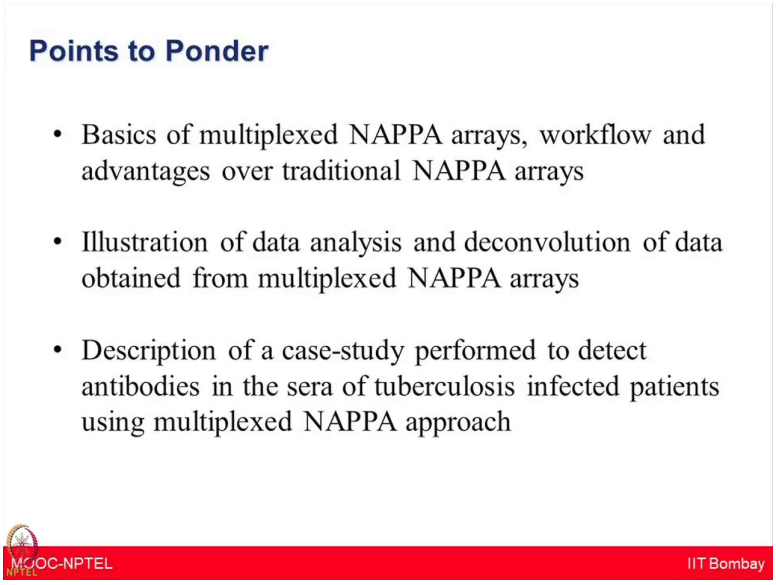
Student: All 40?

Yeah and these results are ELISA results.

Student: So, then it gives a pretty high confident right so?

Yeah, we are pretty high confident with these. But they are 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, there have signal. Now, as you know this is challenging because what we call an a TB negative patient might not be a TB negative patient. aA lot of these patients have had TB before and they do not 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: 20:55)



So, today's lecture you have 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 m TB clones could have been so difficult, but now because Dr. Labaers group has access to all these clones for micro bacterial with 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 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 will continue our discussion about use of NAPPA technology and other protein array platforms especially for the drug screening in the next lecture.

Thank you.