Lecture 5

Biomarkers: Harnessing the immune system for early detection of disease-II

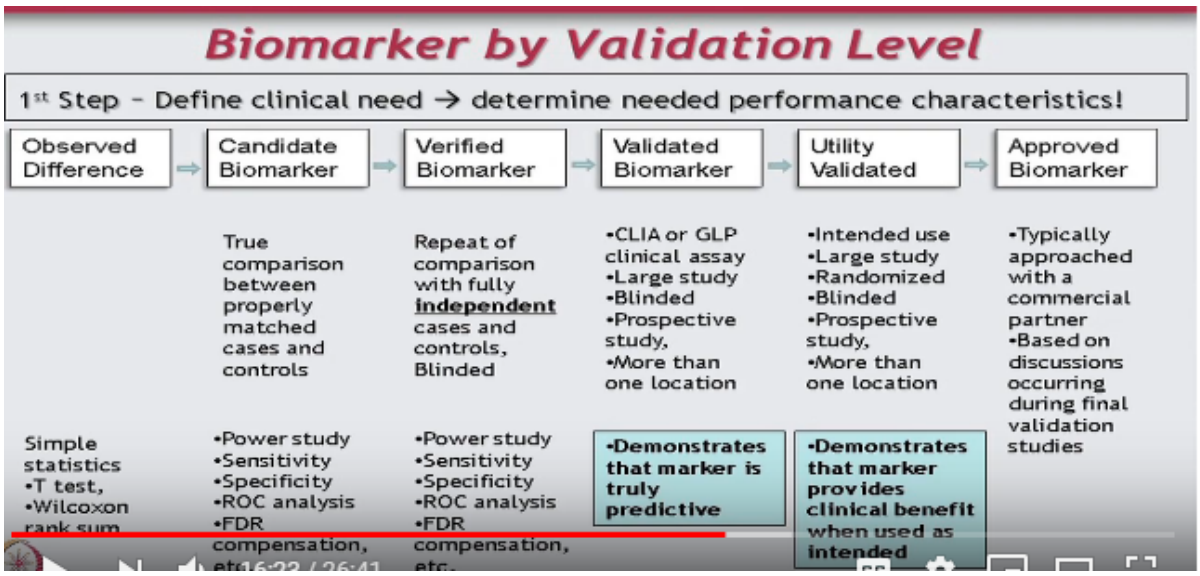
Applications of Interactomics Using

Genomics and Proteomics Technologies

Welcome to MU coach on applications of Interactomics using genomics and proteomics technologies. We have with us a distinguished faculty Dr. Joshua LeBaer, here's another pioneers of the field of protein microarrays especially Nappa technology nucleic acid programmable protein arrays. Dr.

Joshua LeBaer also is the key leader of biomarker discovery programs and he also leads one of the programs operated by the early detection research Network or EDRM, in US, for the biomarker discoveries. So, he brings lot of his expertise. His experience of both technology developments as well as, how it can be applicable for the clinical problems and especially biomarker discoveries. Imagine that you know, we are working like I said earlier, in an Indian contact, I am in Mumbai based and we have the samples coming from Maharashtra, from you know different hospitals from Katta memorial hospital, Kem, Indhuja various local hospitals here. Now our population is very restricted we are talking about people only coming to these hospitals and trying to look at, in a given context of a given disease, what kind of proteins are being changed? And let's imagine that you know that particular protein, a given protein looks pretty interesting, which looks very uniform in a given disease context in this population base. But if you think about can we claim that protein as a biomarker, I think that it may not be the right? Claim right? so, How to claim that you have a good biomarker. Of course a biomarker should be generic, it should be global, and it should really work in you know variety of clinical settings, so, validation is really, really required that is the key for making the success of any potential candidates to the right you know biomarker, for the clinical and therapeutic interventions. Today Dr. Joshua LeBaer, is going to talk to you about some of the details about biomarker and validation strategies okay? So, the first thing that most of us as scientists would do, when we do a biomarker is we would observe a difference.

Refer Slide Time :(2:46)



So, you remember those two graphs I showed you that is the first step you take a bunch of samples you know cancer samples and healthy samples, or early stage and late stage or whatever your comparison is you measure something and you see that the value of that X is much bigger here than here and there's a difference. You say Wow. Okay? The first thing you have to do say I don't have a biomarker yet because. you don't have a biomarker yet but you do have an observed difference and the type of statistics you might do our simple statistics you might do a t-test you might do a little Cox and rank test something simple to confirm that those two values are different. But that is not a marker yet so, now how do you go about getting a marker. So, the next step is you need to say okay? I think I have a biomarker now I need to do a larger scale comparison I have to look at more people right?

And so, we would call that a candidate biomarker and we'll do a comparison between properly matched cases and controls so, how do you match the controls to the cases though age right? Gender right? you are right those are the two big ones I would say maybe as you pointed out the population right so, you are not going to take a bunch of people with HIV in Africa and compare them to a bunch of Americans, who have no HIV that would not be a fair comparison to two very different populations so, you are you need, to you know try to stay within the same community's, same age group, same gender group. Ideally the best matching of cases and controls would be the same group of people that go to the same hospital, except that this group has the disease in this group doesn't, so that's they are coming from the same population and then that's what we call a matched population, sometimes like. In their specialties like in the case of a cancer lung cancer study, you would want to make sure that the cases in they had similar smoking histories right?

Because you don't want to be finding a marker that predicts smoking, you want to find a marker that predicts cancer. so, you have to consider your cases and controls carefully you match them and then the first thing you do is determine how many cases and controls you need to study and how do you do that how do you figure out how many to study? power analysis, yeah, yeah you need that you need to get a statistician to help you do what's called a power analysis and a power analysis is a statistical mathematical study that takes into account how big a difference in the value you expect to see? How prevalent that the disease is? in the population you know. How narrow the variation is in the measurement. That you're making does it vary a lot does it very little it takes a lot of these things into account they do some mathematics and they'll say you know what for the difference you are trying to achieve. You need to do this many people in cases and controls typically when they say the difference you want to achieve though the way they will phrase it is if you want to detect a difference in with eighty present certainty this is how many you have to study and so, you have to deter you have to say at what level you are willing to say I might miss it so, you'll say I'm willing to do this I'll do this study if I can get it 85 present of the time so that's what a power analysis is if you see a study where people are doing biomarkers and they didn't do a power analysis they didn't do it right and I can tell you that 85percent, 90 percent of what comes to my desk as an editor they never did a power analysis right and so, that's a real problem.

All right then you are going to, and then you are going to eventually measure sensitivity specificity. And we're goanna come back to that we're also going to talk a little bit about the receiver operating characteristic curve analysis and false discovery rate compensation but all of these types of mathematics will come into play when you do this first candidate biomarker study okay? So, you did your study and you got it looks promising so, you get a marker and it has let's just said 85% sensitivity, at 95% specificity. So, are you done can you publish. Nah what do you have to do next so, you did a study you, you, you, you did the power analysis. You compared the populations you found a biomarker it has 85 present sensitivity what do you have to do. Well you certainly can look for other people who have done the same work but the simple answer is you have to repeat the study because you are goanna get markers right? typically many of us are going to be studying thousands of variables if I am i array right on the Napa race we have now maybe 15,000 proteins right so, let's say the chance of fight let's say that the if you the p-value that people often say is 0.05 right? 5% so, the chance of finding that value by chance alone is 5% that's what it means when you set a probability a p-value of 0.05 so, take 5% of 15,000 how often am I find a biomarker.

By chance alone quite a bit right just by chance alone when you study a lot of variables you're goanna get you are goanna get a marker that works all right so, the first thing you have to do when you get markers that look promising is test them again on another population and that's what's here so, you repeat the study you verify the marker and it's important in this case to use a completely different

set of patients and controls and that's important why? Why is it important when you do the second study to use different people preventing redundancy okay? May be expand a little bit so, you've already shown that that marker works for that population for whatever reason that marker let's assume you did their study carefully separates cases and controls the question you are asking in this study is. Is that a general factor is that just happen to be a random chance for that one population so, by doing it in a different population you are verifying that in fact it really is for the disease and not just by chance alone so, there's a famous story in proteomics some of you may remember this but at the beginning of this century there was a pot there was a paper published.

In The Lancet it was a proteomics paper and they developed a blood test for ovarian cancer and it was based on mass spectrometry and they predict they claimed that they had 900 present sensitivity and 99 present specificity, astonishing numbers anybody? Who knows anything about biomarkers looked at that and said that's not right? there's no way that you could get a hundred present sensitivity biologist not that predictive well so, they got a lot of press hold programs were started at the NIH around it a huge amount of excitement it was a big deal that proteomics had solved the detection of ovarian cancer and it all failed it was a huge miserable failure and it setback proteomics by a decade because people stopped funding us because they said that we make claims that we can't support and one of the fundamental mistakes that they made in that study was in their validation step they used the same control group they did use different cases but they used the same controls and so, they didn't follow the rule that this group has to be different from that group and consequently.

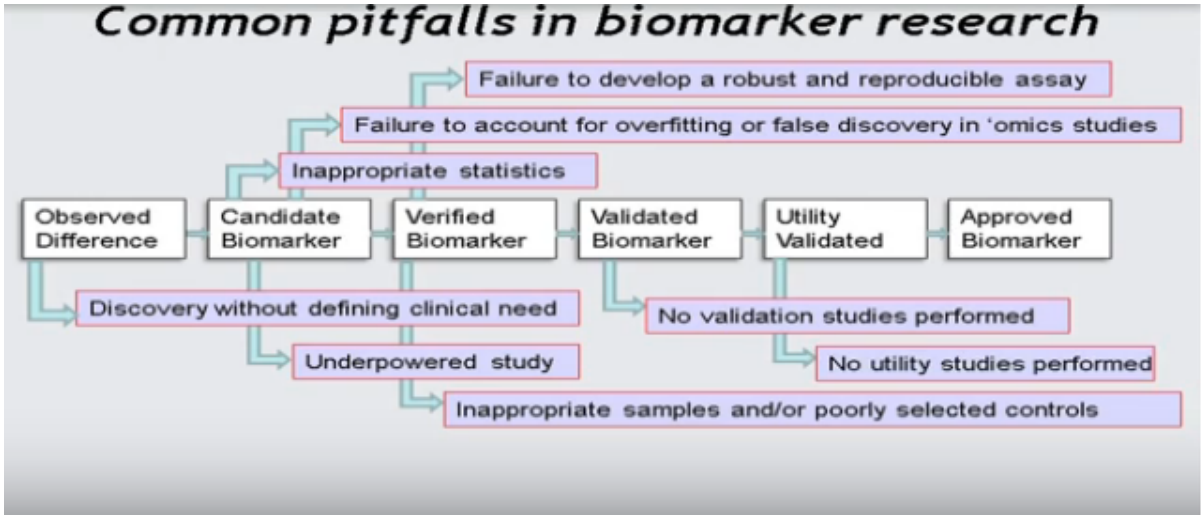
For whatever reason that control group had a defined pattern that was definable as roll and that's what allowed their biomarker to work but it was just random chance it had nothing could do with ovarian cancer and so, that was a huge error so, you have to be careful about that so, that if you get to this point and your marker still holds up now I think you are ready to publish at this point you can say I've got a verified biomarker this is worth telling the world about and. And then you can send it out for you send it out for review I will tell you as an editor for JPR if I don't see this I don't even review it I send it right back to the author if they don't do a validation study they're out I just I won't even look at it all right so, then I'm sorry. it can be as long as, they are different people. Different people they can be from the same Hospital they have to be different controls, and they have to be different cases oh there can be no overlap and the people no same disease. But different people yeah so, for example if let's say you. You have a you have 200 people, with ovarian cancer at your hospital and you found 200 women, with you know that are good controls you could split them into a hundred cases and 100 controls, and do your first study and take the second hundred and the second hundred and do your verification study that be perfectly good design okay? So, after you get your verification you still have a long way to go to get a valid marker now you have to do what is called validated biomarkers. Study these sorts of studies are typically a level past most academic labs most of us can't do these studies they have to be done very formally these studies should be done under what is called either CLIA. Or good laboratory practices certification.

They should be large studies they should be blinded studies blinded means that the. The scientists who are measuring the values do not know who has the disease and who doesn't write and. And all of that is hidden in the documents they have to make their predictions based on what they set back here typically there should be a prospective study what do I mean by prospective study right so what does that mean that's right so you're not looking at samples that you collected last year you're collecting samples in the same manner that you would be doing it if you were treating patients you click the sample and you test it and you see whether it predicted properly or not and then you you need to do these in more than one location if you get this done what that tells you is your marker is truly predictive it really does predict the disease that's great that is already something to be very proud of

now you have a marker that predicts disease are you done. You can tell you are not done because there is still space left on the slide right. right? So, there's still more to do right so, just because the marker tells you that the patient that can predict a disease you still don't know if it will be a clinical benefit using that marker and so, the next step you have to do is what is called a utility study you have to ask if I use this marker on a population. Will it tell me something that reduces mortality or morbidity in that population because?

I detected the disease early okay? And so, here what you do is the same thing as here randomized blinded study prospective study but in this case you are doing it as an intended use you are measuring, you are measuring it you are predicting an outcome you are telling the patient and you are acting on the prediction and you ask the question in those people with whom I use the marker did they have a better outcome than the people. Who did not use the marker did the marker save lives did the marker reduce disease and this is where a lot of markers fail so, some of you may be familiar with this marker called ca-125. Which is a very good marker for ovarian cancer there is no doubt that ca-125? levels correlate with ovarian cancer that marker is used all the time as a disease progression marker to monitor ovarian cancer it's quite specific the problem is if you do ca-125 to detect cancer it you don't see any better outcomes and the. The problem appears to be that by the time the ca-125 levels are measurable the cancer it's already too late. They it doesn't come up early enough and so, it is it is a predictive marker so, it fits it succeeds here and it fails here if this works here then you get an approved marker and now you are in good shape I can tell you that this whole process is very long very expensive and has only been successfully done a handful of times okay? So, what are the skill sets that you need to accomplish all these tasks right.

Refer Slide Time :(16:40)



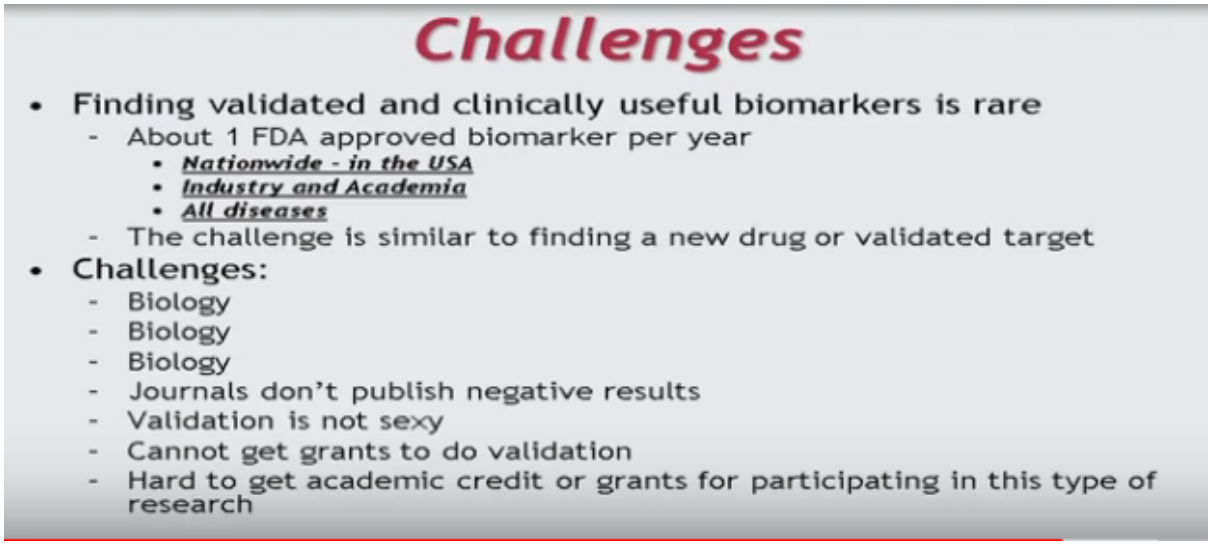
And so, that is what is shown here and this is just to emphasize. But to get a good marker you need a multidisciplinary team there is no way around that so, you need to have early on you need to do these first sorts of studies you need people with molecular and cellular biology experience. Throughout the study but especially at the beginning you'll need genomics and informatics as you go further into the study you need good statistics. You need to develop strong robust markers that you can that do in the clinic you need good analytical chemistry obviously you need good clinical understanding, and understanding of epidemiology and then when to, when to use these markers depends on looking at health policy so, at different stages of the game you are going to need different experts but throughout the whole process you are going to need a lot of experts okay? Yeah, well how you do it varies a little

bit it depends the way you do the power analysis depends on the study and what goes into it so for example oftentimes when you are at this phase you might be doing protein arrays or you know next-gene sequencing or some kind of large-scale omits scale study where the number of variables is very large and the type of power analysis you have to do with large variable numbers is different than if you're testing doing a power analysis for just one marker that you have as a predictor.

In this case you may have to do modelling statistics to get a good predictor you might have a simple formula you could use over here but the idea is the same it's just the execution is different okay? So, where that where does this go wrong so this can go wrong in a lot of places and it does all the time so, the first mistake is you, you, you discover some kind of a difference but without defining a clinical need you haven't defined the clinical need your difference may be meaningless or may be useless. People often do inappropriate statistics on these candidate biomarkers they'll look for p-values. Instead of doing proper biomarker statistics people don't do and they do what's called an underpowered study what's an underpowered study .The what the sample size is too small, yeah exactly the sample size is too small and there's two consequences to that the first consequence to that the most common and historical consequence is that if you. your sample size is too small then you, you run the risk of missing a good marker because you didn't study enough people you didn't you, you**,** You, you, you, you won't have as enough of a chance to find the marker in the modern era, the problem is a little bit different the modern era the problem is these days we don't study a few variables we study tens of thousands of variables and so, in the modern era an Underpowered study usually means that you are goanna find differences that are meaningless you are goanna find random chances that this gene is different from that from in the cases and controls and it's not really to the cancer at all because. Of what is called over fitting, over fitting is statistically finding something that isn't really real and it's a huge problem in our field I can pretty much guarantee you if you see a paper published and typically they are published in the best journals science cell nature you'll see a paper published next week a month from now on a marker. That has 100% sensitivity, and 99% specificity, and if you look carefully they probably over fit because no markers are ever that good okay? so, failure to account for over fitting I just said it you heard it here first using inappropriate samples poorly selected controls so, people don't carefully match the controls so, for example I've seen studies where people used a bunch of cases from one location and then they ordered their controls from a company and then they compared the two and sent I found a marker I can tell the difference. Well they can tell the difference between samples that came from the company and samples that came from this hospital they didn't necessarily find the disease impact of you know Paul temps he is a proteomics researcher at Sloan-Kettering Paul did a study where he was looking at prostate cancer he was trying to replicate the kind of approach. In that ovarian cancer study I mentioned earlier that totally failed but he was trying to do it right and what he found was he was looking at State cancer and he found a marker that was remarkably good at predicting prostate cancer.

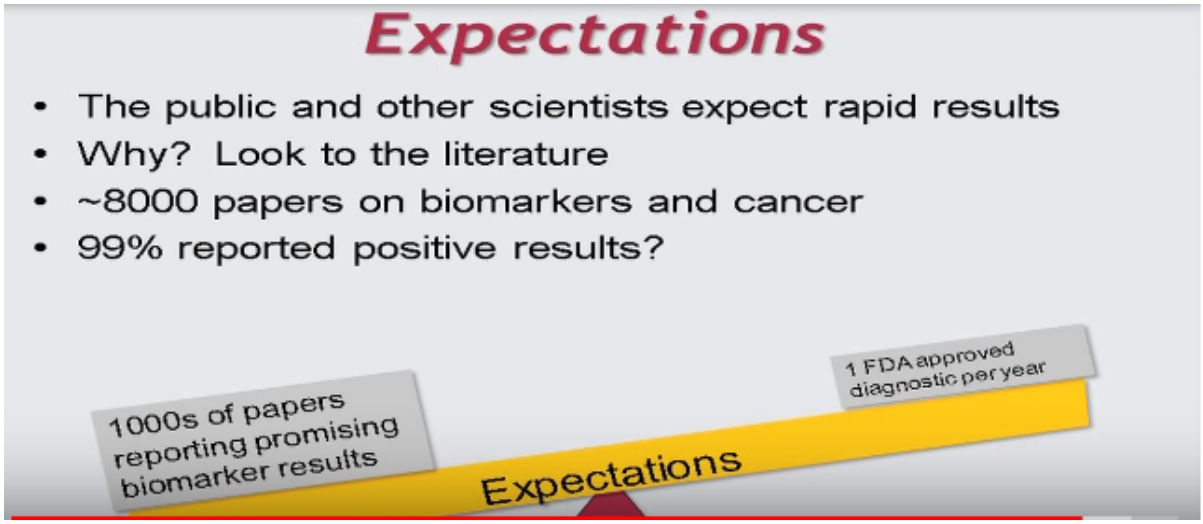
But you know credits to Paul temps because he, he looked a little harder and what he realized was that the prostate cancer samples .Were all drawn in blood samples, from men who were about to get biopsies they were all in the hospital and they were goanna get biopsies and the samples that came from the controls all came from the outpatient clinic and it turned out that the two, two locations used a different manufacturer of the blood tubes so, the blood tube type was a little bit different and when he did all the analysis what it turned out was he had found a really good biomarker. For blood tube type and nothing to do with the disease at all it had to do with the types of the tubes that it came with so, you have to be very careful. And so, and then people often fail to develop a good robust and reproducible assay if you are goanna do the kind of late stage validation here you need to have a good assay for that some many people forget to do this study here or they don't do this study here.

Refer Slide Time :(22:41)



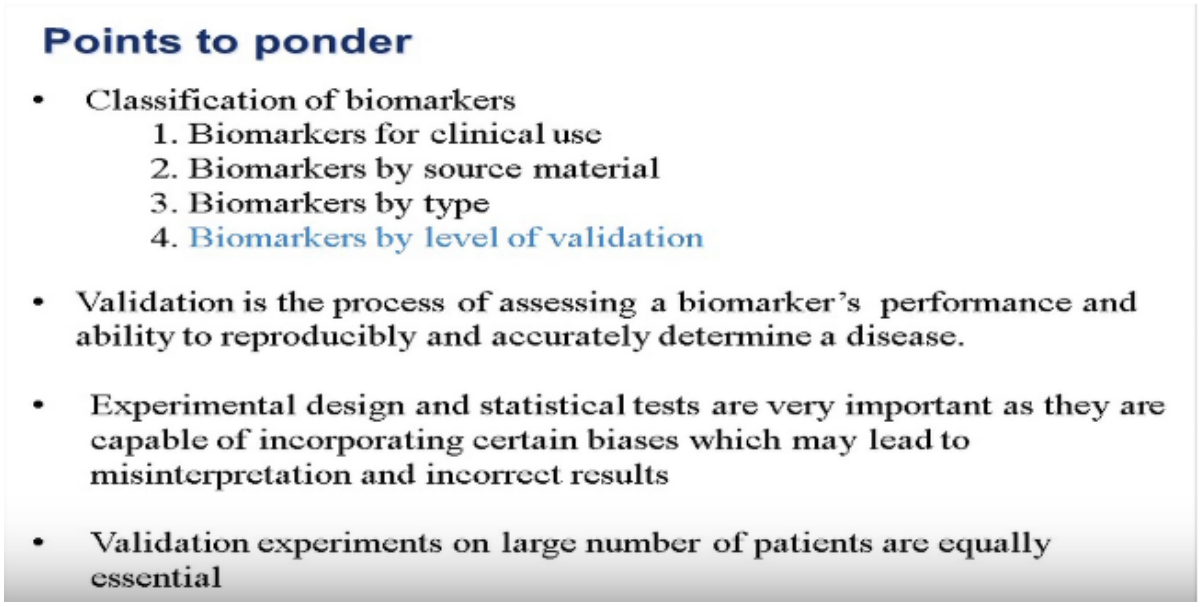
And so, that that's that sort of summarizes some of the major problems that that you can encounter Okay? So, lots of challenges finding a good clinical useful biomarker is very rare these days in these upon average maybe one to two biomarkers a year. Will succeed in making it through the FDA so, this is very, very challenging and that's combining all the work of academia and industry all combined that's all we get I would argue that the biggest chain the big biggest challenges are the biology itself it's very hard to find a molecule that specifically can predict the outcome of a patient so, you have to look extremely hard to find it but journals don't publish negative results and so, oftentimes people don't realize when, when markers are bad and so, they end up you know only publishing bad biomarkers no one likes, to do validation in fact in and in in the us it is very hard to get funded to do a validation study so, let's say you do a good biomarker you have all the best intentions you do the observe, you do the observe difference you do the initial study and then you do the verification study and you say okay? Now I want to validate this marker the response you'll get on your grant application almost always is well you've already studied this marker why do you want to study it again. And you can say because I want to validate it they're like no not you already studied it you're done no I'm not done.

Refer Slide Time :(24:12)



so that's exactly one of the problems that we face all the time all right so, let me move on then so nonetheless the public really expects to see these results and that's partly because there's thousands of papers that report Good bar biomarkers and there's usually only one good one per year and so, everybody thinks that it's easy but in fact it's really hard so, that's kind of a take-home message.

Refer Slide Time :(24:37)



All right so, just to conclude Dr. Josh River has talked to you about different basic consideration How you can be confident that leads which you've identified as a you know potential protein Candidates whether you can dump them as a biomarker. What type of test you should do both from the Statistic point of view as well as the right clinical assays in the clinics in the labs which can ensure that the candidate. Which you are identified that is actually a potential biomarker so, these basics are very important for you even if you are a student, or you are a researcher, who are planning to be involved in the biomarker based programs I think you know your strategies thinking about the power calculation these statistics looking at this sensitivity and the three of D biomarkers as well as your plan to do validation of the candidates becomes very crucial. I hope these basics are really giving you new insights about how to now utilize this understanding this knowledge for the Actual clinical applications.

Thank you very much.