Introduction to Proteogenomics

Dr. Sanjeeva Srivastava
Dr. Henry Rodriguez

Department of Biosciences and Bioengineering
National Cancer Institute
Indian Institute of Technology, Bombay

Lecture – 60
Proteogenomics: Opportunities and Challenges

Welcome to MOOC course on Introduction to Proteogenomics. Today, we have last lecture of this whole course. In this MOOC course, you have been introduced to the concepts of genomics, proteomics and proteogenomics, and effort has been made to help you understand various steps of data generation, analysis and interpretation. Though the field of proteomics is a still evolving, its contribution to the development of science particularly in precision medicine cannot be undermined. There many tools which are currently being used for proteogenomics.

I hope you got a good understanding of these tools and publicly available resources which you can also start using for your own research. The National Cancer Institute, USA has constantly made an effort to bring research communities together for fighting the common evil of dreadful disease like cancer.

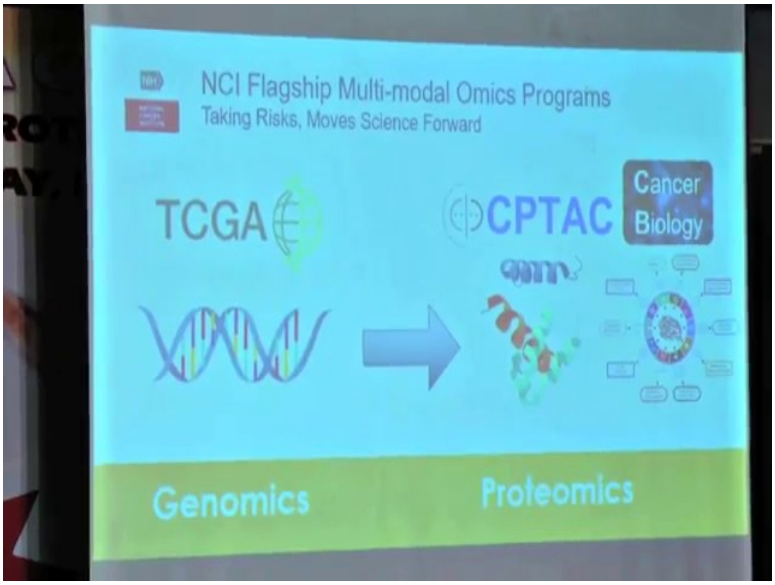
In this last lecture, you will be introduced to the various initiatives of NCI toward development of a cancer free world. This lecture is essentially a brainstorming meet of cancer clinicians, researchers and industry experts which we conducted to mark cancer moonshot India Program at IIT, Bombay. So, let us have this interactive session about cancer moonshot India and a perspective shared by Dr. Henry Rodriguez.

Thank you. It is a great honor to be here, I mean I know, but we really have to go, when I went to Sanjeeva. Sanjeeva talked about following use the developing in India. The one of the things I am doing was contacted a lot of my colleagues that knew his work and the one thing that I did is that the work that is being produced in this laboratory is exceptionally well.

So, it is just an honor to now to be here knowing that Indian are joins international effort, there is one of the out shoots of the United States Government. So, what I thought that I would do is to sort of give a simplistic overview of how we ended up doing what we do now

within the national cancer institute, more specifically 2 years ago how the cancer from shot actually got established.

(Refer Slide Time: 02:52)



So, let me get my little history here from a simplistic level. People do not know why that national cancer institute enjoys these large health programmes. So, I actually do not know what proteomics background who know me you can look up, I actually looked up genomics background and classically trained in California and drug development.

So, what is the things that did not attract me to the national cancer institute, they recruited me was their history and the history needs to understand what genomics did to proteomics. Though if people talk about genomics large scale and you mentioned the word national cancer institute that everybody is going to recognize is a cancer genome atlas and that program actually got created in 2006 when it went public that is the key when a program went public. And the program in a 10 year window has done an incredible job for my cataloging please perspective to develop these great resources that the public is able to use.

In this span 10 years, they went through about 35 different cancer types looking and in catalog over 14,000 individuals. But the part that a lot of people might not be familiar with this that when they want the cancer genome atlas or when they are trying to develop it they did not simply want to go after the genes, the national cancer institute they were formulating this program they wanted to go after proteins and at the very same time that we launched the

TCGA program based on the genomics landscape we went after proteins. And then program is one that affectionately referred to as its CPTAC.

Now, the reason they wanted to go after proteins at the same time as genomics was for two basic reasons which we are talked in the various sessions that people now have been holding. One of them is you absolutely want to figure out the biology of cancer I am one of those people those think that Biomarker discovery is really great, but unless you understand the biology of the disease. It is very difficult to keep a novel discovery that you find which is an anecdotal observation and making that discovery clinically actionable on a wide scale that is very difficult and very rare to be quite honest. So, understand that biology is extremely important.

The other reason that is very important to understand the protein is exactly the what they are really going after is the keyword therapy, while the immuno-oncology is very promising the vast majority of drugs that we still give to our patients, they are typically chemical base and the chemicals are very few that target DNA such as inter binding strands. The main variety of drugs will target a protein. So, you really need to understand from those perspective what is the quality of these proteins; and exactly the efficiency and the binding constant on the target you are trying to go after; not an inference which is typically commonly done.

But here is now what happened. Before the cancer genome atlas got launched, we were starting to think about it in the early 2000s and that is when the first draft of the genome got created, that really led to this great interest on looking at the molecular biology of cancer. But at the same time there was a publication that got released looking at ovarian cancer early stage using an emerging technology of the time which was mass spectrometry. They make it a claim the drug that they can use proteomics without even recognizing the protein and simply recognizing a pattern of an instrument and using that as a predictor for early stage of a cancer. Very promising and raised a lot of interest with a lot of cancer directors back in US. But fortunately it was found that the study, the way it was designed, the way that things were interpreted were not correct.

So, the reason what we headed at the end say was quite interesting, when it came to CPTAC with network of the biology in 2006. Unlike the genomics landscape they felt at the time the technology was quite mature and you can trust the data. So, CPTAC first have to show that you can take these emerging technologies and do your best to try to understanding analytics,

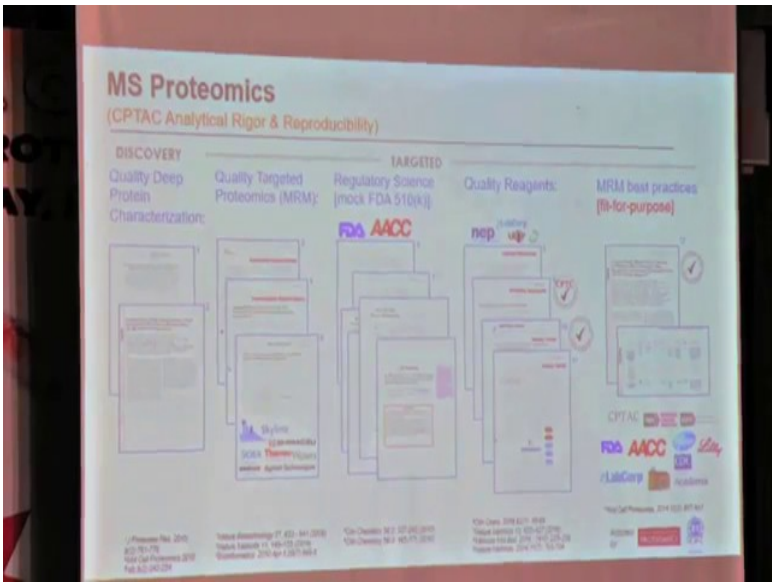
standardize where you can, and if you cannot standardize, try to harmonize the measurements in the analogue workflows.

Once you are able to show that you can actually come back with measurements is going to be representative of biology, not the measurement that is representative of an artifact that the we take a sample, we process a sample, the way you do your instruments, then the world give us permission to work the biology.

So, this is what we ended up doing which is quite interesting because it is very rare that actually to help develop the standard tradition. We ended up doing them to bring proteomics to the state of genomics. So, for the very first 5 years CPTAC basically try to go after the analytics of mass spectrometry and we look at two poles which actually discussed in the past two days.

One, we looked at the discoveries things. Here, we basically showing if we take a lot of people refer to my shotgun, I quite frankly this is unfair to the terminology shotgun, but I intend to basically refer to genomics, this is a deep dive comprehensive measurement of trying to look at everything we can assemble, exactly we do in genomics.

(Refer Slide Time: 07:54)



And in that space we basically show need to get standard operating procedures, we did elaborate problem with study, we actually have any laboratories throughout the US and we

did an international front and we sure he gave very good concordance of your measurements across your laboratories.

But, the one that we really wanted to put our mark on this is the one that exactly we do in genomics. Once you do a deep dive they need to develop gene panels. These gene panels is actually that drives our clinical trials today. So, we wanted to develop that same space when came the proteomics.

Now, it turns out this measurement technology a lot of people refer to it today has multiple reaction monitoring, you have different ways of crazy phrases, but this is not something that that CPTAC development and attention imagination. It is being used in clinical lab for over 30 years. He simply used it for the measurement of small molecules, but we wanted to shows other laboratory were already using is gold measurement of targeted mass spec imply to the measurement of of the peptide. Informatics, he speaks the information back up to the measurement of a protein.

So, we end up doing there we basically look at MRM, this is already being used in back end laboratories, but it is really due diligence on the accuracy and precision across multiple apps was yet not demonstrated. It is very important that have that proof before you go back to tumor board. So, we basically get with lot more studies we have laboratories to the distributed one of the US. Then we did it in international forum that can we show, that this is a very good quantitative reproducible, reproducible measurement tool.

The other thing was that we wanted to do to time was explored the clinical space. If you find an interesting biology and if the biology is best measured using these technologies what would it take to get the technology approved by the regulatory agency? So, we ended up doing one of the things that we did was quite nice that in the US to get a diagnostic device approved such as an IDMIA and need to get regulatory clearance.

And almost there is two types, the first one go and look frequently 5-10 k. So, we actually worked with the regulatory agency and states, we worked with the clinical community, more we ended up those quite normal. Typically, manufacture most of them 5-10 k to the regulatory agency and the regulatory, then the FDA will mark up the documents. Bring all your comments to concern on what they just submitted.

But typically, with that goes back to our company though there would be release it to the public. We decide we wanted to make that very transparent. So, we worked with them we put up on a workshop and we actually submitted, get official filing with the regulatory agency using this type of the measurement technique. But, because we made up all the data, but we did not make up our analytical workflows, it allowed the FDA to market the document. And then once we got the documents back because we worked with the chemistry community we published all their markings up. Such a great way of making very transparent exactly the kind of questions you would give, if you would just make your instrument where these measurement techniques they get them approved by the FDA.

Here it recognizes that a lot of the reagents were being commercially sold we felt that the quality was not the local standards that we wanted to see these within the research and all where it can be within the clinical great work. So, we worked with various manufacturers elevated these standards when you saw these agents in the public domain.

The other one was we started going to meetings of people would always saying, I have a assay, I have an assay. In fact, one millionaire that I went to that quite openly is that one person stood up and they said I already got announcing that every human protein that is out there. I was quite surprised when I heard that. After the meeting I basically approached this individual and said explain to me how you developed that assay that every human protein, where turns out what they talked about was a theoretical based assay, when an assay is basically running on buffer. In a clinic that is not considered an assay we can use that terminology.

So, we decided to do with CPTAC that we basically then started develop fit for purpose these criteria that begins to define exactly what an assay is. What is nice now is that that has not been accepted by the international community more specifically one of the prominent journals of FCP. Now, they have adopted those criteria within the journal itself. And this was also done with the pharmaceutical industry with accredited laboratories and regulatory to see, and the clinical labs in the United States to develop these sort of criterions. So, this actually now represents 5 years worth of history with the CPTAC based effort.

Again, we had to go back we had to show that the measurement you are able to obtain is basically trustworthy you can actually believe in the measurement people representative biology. Once we did that, we went back to report and then we got the issue. And we decided

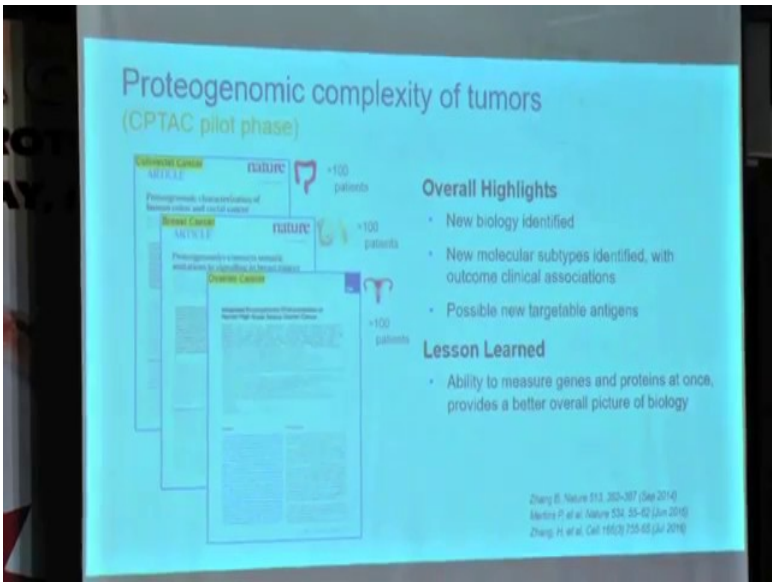
to do is quite interesting. Because obviously, at the NCI we have the cancer genome atlas, that now has 5 years with the history of our CPTAC. They are generating a lot of interesting information and what our proposal to them, was we want to take the exact tumor, that just got genomically sequence within the cancer genome atlas program.

And at the same time that we would actually now go after the proteins within that sample. And we will believe that a unit of layer, a comprehensive protein than about a comprehensive genomics measurement you are able to obtain additional biology, that is either difficult to obtain or simply not feasible through genomic itself.

Now, think about that, because at the end of the day and I saw today a lot of people say, oh, proteomics that is always much better than genomics. The reality is when the company move to the clinic two things is you can drive your decision, and that is, is the test clinically relevant for the disease to try to go after and another thing I can ask is for how much is the test and what is your throughput. Because the reality is if transcript omics is able to predict the same thing for proteomics can, so many that the hospitals going to say why would, why do I need measure in my proteins because it is more throughput and it is a higher cost. So, that was the cable that we talked could we find additional biology.

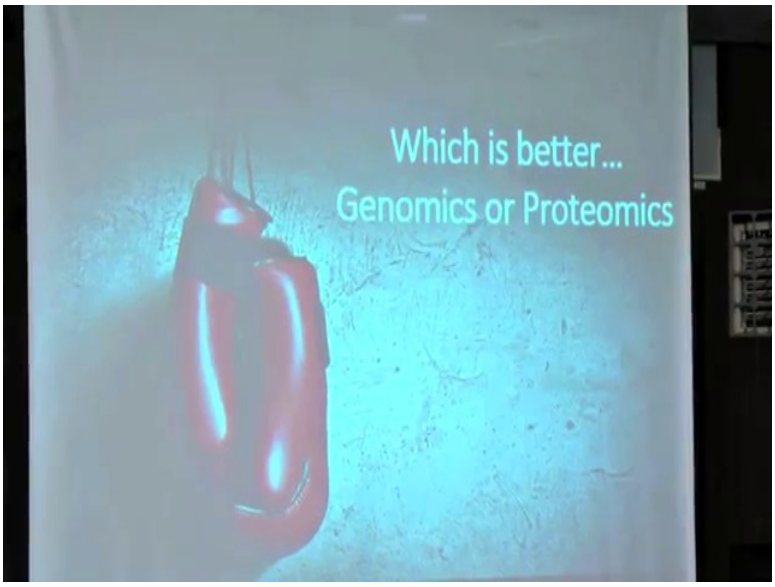
So, in the next 5 years we decided to go up to 3 cancer types in the cancer genome atlas. We would have the breast cancer, ovarian cancer and colorectal cancer. I am not going to go through all the details, but suffice to say in each one of these cancer types. We were able to identify additional biology that was missed simply because we can obtain it from genomics itself or it is just a better way of integrating the data set between at least within the genomics and within the proteomics landscape.

(Refer Slide Time: 14:46)



Furthermore, what we learned from this lesson was that if you simply go after one type of involvement whether be genomics or transcriptomics or proteomics, most likely you are going to be missing key biology that could be infirmed from one of those other omics. So, integrating those worlds would become very important for our program. So, with that in mind I heard that question today so I quickly did not put this slide in.

(Refer Slide Time: 15:25)

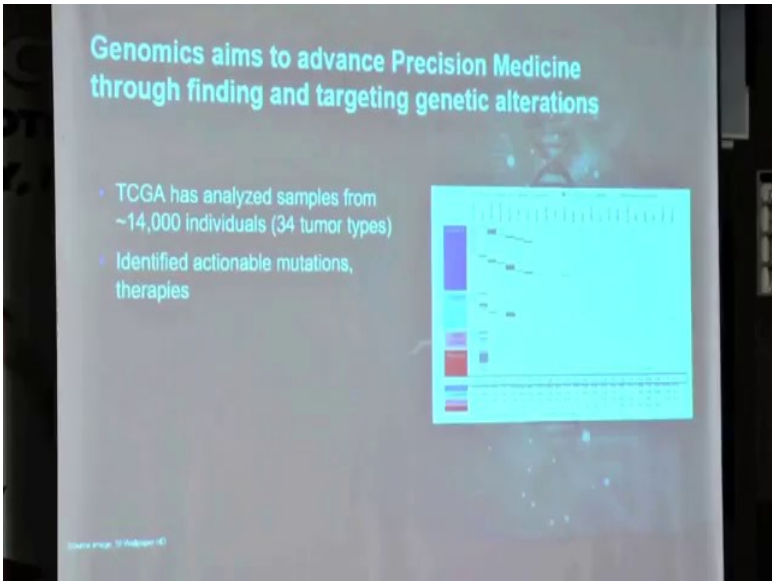


One of the things that I started to be if that people were asking me, in fact, before I had not went back to my board, to come up with the next version CPTAC was the exact question that

somebody just asked at today's conference and that is well you take up proteomics and throw up on genomics. And you can expend all this additional biology, why do not just do proteomics then, this much better. Well, my philosophy has been should you do genomics and proteomics and which ones better, I would argue no one really knows the answer definitely.

And here is why, I will break it down into components for you. One, let us look at biology itself because that is the part that I tend to love the most. If you look at the cancer genome atlas because the TCGA what they did for 10 years it really using clinically actual actually deployed well. They are basically trying to figure out biology as those samples were not collected with a clinical question get involved, it was basically cataloging samples.

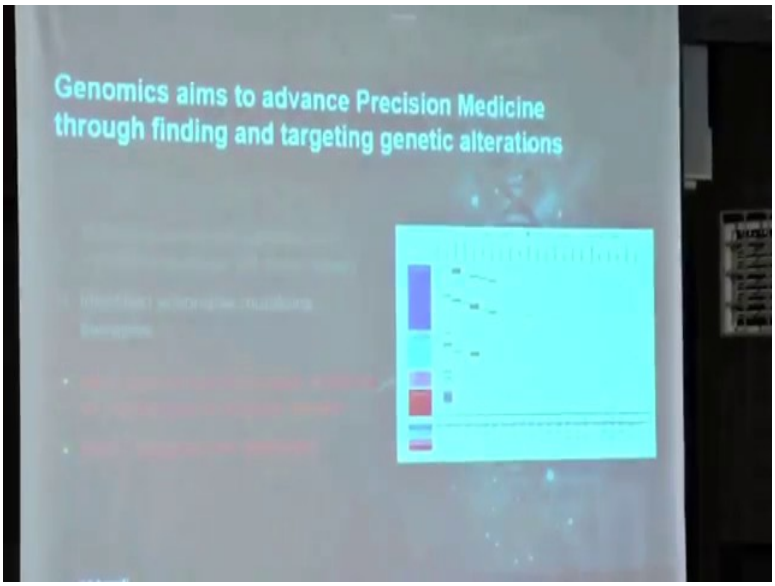
(Refer Slide Time: 16:16)



TCGA as I said in a 10 year window, they would have covered 3-4 cancer types and they basically genomically characterized over 14,000 individuals. In the process, they basically identify all these action orientations. We now we have small molecules, all these small molecules are driving allow for precision oncology trials. So, that is the good news.

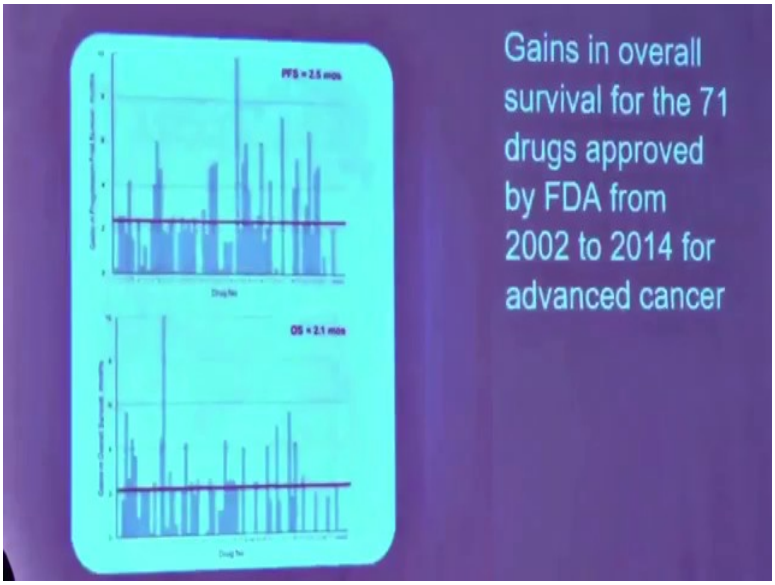
Now, you can look at the other side of the story which is who were learning about 3-4 years now into this sort of science and driving with our clinical trial. More were learning that a lot of these individuals that we identify all these actionable mutations, a lot of them really are not responding that not well to the treatment that they are being administered within our treatment aims of whatever clinical trial that they are being put on them.

(Refer Slide Time: 16:52)



Furthermore, that they do respond, but find out that those responses are actually short term to live and that you can exclude toxicity that could occur that once you get that you have to put them on another or more drug trial itself. So, the question becomes, why? We do not know. But what we do know with absolutely certainty is that there is still a tremendous amount of biology that is missing from that picture. Now that, now this is the biological version, let us flip it to a clinical way of thinking.

(Refer Slide Time: 17:34)



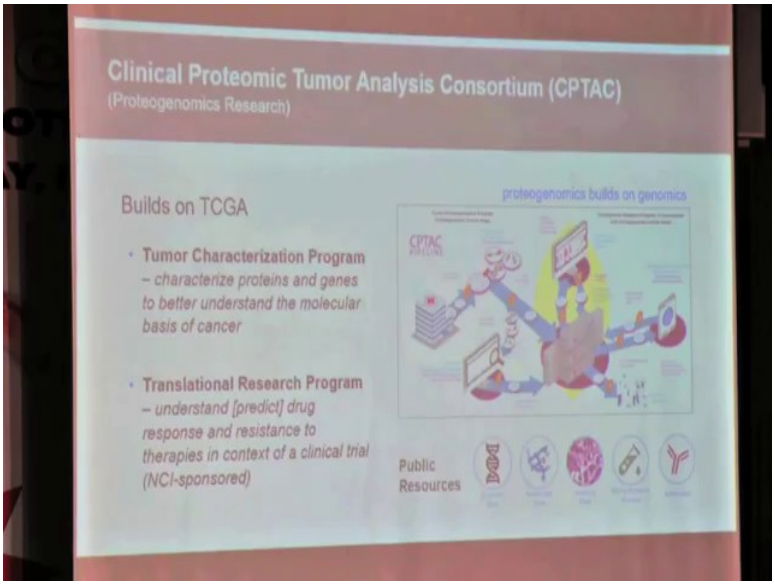
A great little paper came out about 3 years ago. And this is by Tito Poleward, used to be at the cancer institute in Nigeria city. So, Tito when he had doing its quite interesting, he basically said let me actually go through the regulatory agency back in United States and asked a question about a 12 year we know people study to determine, medicine precision, oncology, the reality we are talking targeted therapy.

And look at all the targeted therapy drugs that now have been proved by the FDA that's just, these are about over 7 now, and your window spans about 12 to 15 years, and when he had doing was he took two common criteria's, it is used all the time in our trials. And that is when you look at individuals to ask well what is going to be their overall survival and at the same time what is the progression free survival of these individuals.

He excluded the exceptional responders, which is what a lot of people love to go after and again to be very fair this is solid tumor typically in human stage. And when you follow down ones for all those drugs now that that have been approved from the targeted perspective on average those two criteria's its less than 3 months.

So, and that is really not that good. Again, it is very promising in oncology, but it is still we could do a lot better. So, using these two criteria's, the argument of missing biology and then the argument of can you begin to play in the sand box of clinical trials that influence directly the next iteration of the CPTAC program.

(Refer Slide Time: 19:06)



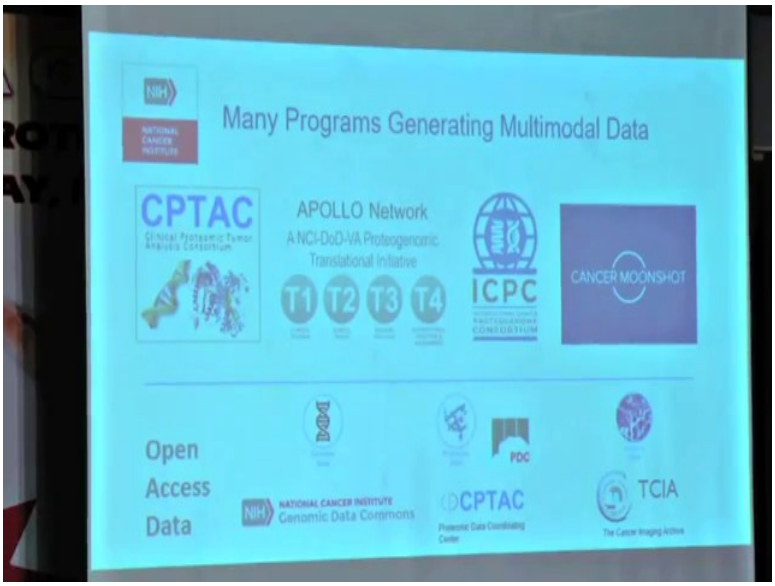
So, this is actually CPTAC. So, this is actually CPTAC today. Again, these are 5 year programs and the next iteration moves without issue that is one, but that would be contingent on the science that comes out of these programs. So, we still have a biological law that is exactly what we did within the next component, were being held response to go after at least 5 more cancer types, hopefully when we work, but it is a minimum we have to go after 5. Every sample that we get from our patients these are all treatment or AID samples, every sample goes every partner directly with a cancer genome atlas. They will do comprehensive genomic characterization then the sample of pieces are also goes to our characterization centers along with our data analysis centers.

At the same time that we run a biological arm. We have now an official translation arm. For the very first time, the cancer research is now partnering an ongoing proteomic laboratory with an NCI sponsor of clinical trial. With the 3-types of cancer types that were going after, after over that component that involves a series of drug trials.

But again, the part that I think that is quite nice about CPTAC is that the data we got born to about 12 years ago now, everything that we produce we put it in the public domain it is listed on the bottom of slide. Everything from genomic information, proteomic information and your reagent that we develop which are typically nullified any of these assays all they are so pleased that we produced our assets against, all that is placed in the public domain. The argument is that we know its suddenly being used by the community, and we believe by giving it back to the people it drives the science and hopefully patient care you are able to accelerate not, I just living your country, but across the globe is that our goal all of them.

So, once this program got launched, here is something now that is, so here is now what is going happen during the cancer institute. So, CPTAC, while one of the first ones to start to mix these two worlds together from a programmatic official level, it is not the only one that is going to do that now.

(Refer Slide Time: 21:14)



There is two other ones that recently got merged just over 2 years, it will be to the cancer moonshot, whatever is referred to as the Apollo program and another one is referred to as ICPC which is the International Cancer Proteogenome Consortium.

(Refer Slide Time: 21:30)

So, very simplistically the Apollo program is one. So, these programs actually got started as follows. I think a lot of us and especially myself I was extremely moved in the early 2000 since by the inspiration call of the former vice president on country Joe Biden when he actually wants the cancer moonshot.

Now, you go through all the details of what the cancer moonshot is, but the part that I enjoyed the most was I tried to simplify it into 3 simple objectives that we want to achieve from them. Whatever happens to be, we want to accelerate the progression in cancer research. There is many ways that you could do that, you can do technology development and other components.

But, the other two were the ones that I have and Macaron have always which is in CPTAC has been doing from past 10 years. One is, wanted to see greater cooperation in collaboration and to be very clear the way it was phrased it was not winning your own university, it is not winning your own country, they are hoping tried to explore international of a collaborations.

And the third one is the one that I was very happy with they wanted to see a lot more sharing of your data. The reality you can look at the genomics landscape, a lot of the information that you are developing is pretty much pretty competitive because they are basically observations and a lot of those observations typically does not have yet a clinical relevance behind them. So, releasing that data is not detrimental it a fact it is actually beneficial because other individuals are able to take your data set, you get recognized for, but again you are able to try science a lot further.

So, using that as a backdrop the very first one that we wanted the pilot, taking the CPTAC model and trying to bring other organizations into it became one that involved the US Federal Government, specifically 3 of our agencies and that program now is referring to as Apollo.

(Refer Slide Time: 23:23)

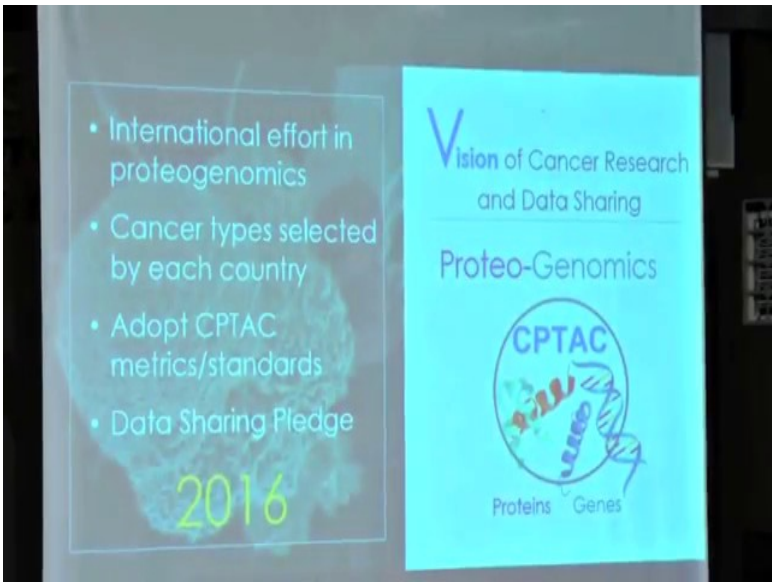
So, Apollo is one right across my hospital have to be one of the largest hospitals from the department of defense which is referred as cancer center. If I that is the one that we see a lot of the presidential helicopters and congress, and all the representatives get treated out, but that said though we basically walked across our street, we met the cancer center director, had great conversation, what we recognized is there was an opportunity to begin the pair of the National Cancer Institute, the department of Defense and the various administration.

And the game is very simple. They will begin to adopt a lot of the metrics, a lot of the standard developed by CPTAC and they would also begin the implement this sort of proteogenomics based approach to look at the science of their veterans and other family members.

And the part just quite nice again is that we scientist partnership based on dumpling one main criteria is that all the data that they would produce would be placed in the public domain. So, that is what that involves the US federal government. Then the next thing I find out what I am called by the white house by representatives about the cancer moonshot program they said, so we kind of like we ended up doing from this federal government perspective is your opportunity to basically make this on an international level. I started to think about it and they deal was quite appealing have to admit. So, this is what we ended up doing.

So, we started to ask myself the following, what if you can actually think its proteogenomic model and begin the scale it on an international level. If you were to do that they would have each country which is the best in making these decisions along with your various representative government, they would be in the best position to determine what cancer type would be a most significance to their own nation.

(Refer Slide Time: 25:14)



Furthermore, they would adopt a lot of the metrics and standards if applicable developed by the US CPTAC program. But again, part that becomes very critical for me, happens is that of any data that you produce from any of these official partnerships, regardless of where you from what country we want to see the data be placed in the public domain and posted in your country. But, at the same time the United States, the cancer institute we would host those data sets for you. That becomes a key criteria what decided our partnerships.

So, that said here is now what change powers. So, in probably 2016 in midsummer, so in January the cancer moonshot gets launched, then we quickly launched Apollo and then I hit this call we would like to see the scale on an international level. The very first country that signs on in July is one. So, at that point we bring in Australia. So, we have 4 institutions now within Australia that well disclose the partnership. I thought my job was done. I give myself one nice pats on the back, I got won, I have been to the white house, I basically go home and tell my daughter, oh crazy stuff I am doing.

Next thing I get another phone call a week later. We loved what you did, is your opportunity to expand this. And, by the way we want to see it expanded in 8 weeks. Now, I had no idea why 8 weeks was important. It turned out the reason it was I found later on that something was happening in New York city at the United Nations, but they were not telling to me at that time. But, I thought that do not be something very interesting to go after. After all it is very

rarely if you ever get the word from the great leaders within your own country. So, I thought it would be interesting. So, we did.

(Refer Slide Time: 27:00)

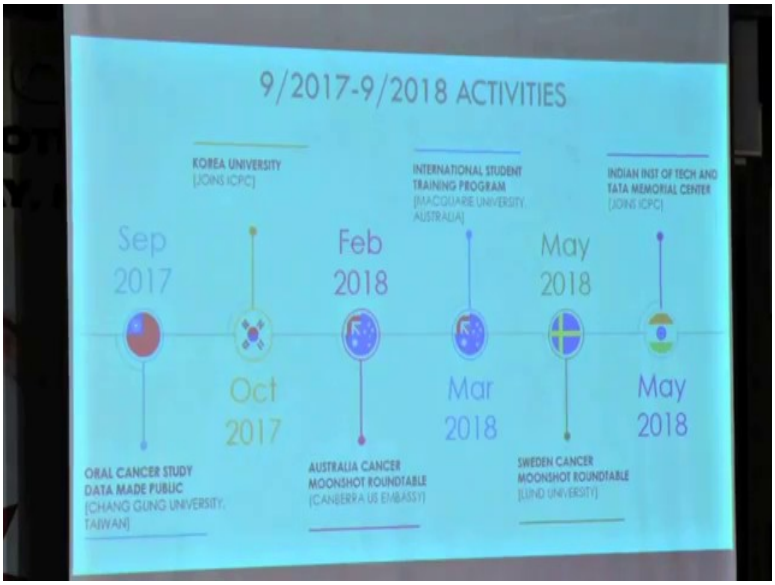
So, in the span of 8 weeks we went from one country, 4 institutions, at this point now we are spanning 8 countries and now its involving 16 institution pretty impressive. But it is amazing when you send an email and it says on behalf of the vice president of the country, this is something we would love to see happen, it is amazing because my name does not carry much weight, other people names do.

Now, this is September of 2016, keep that in mind. Right now pretty much at the end of 2018, so the question is whatever happened to this program, so this program actually has taken a beautiful life on its own.

(Refer Slide Time: 27:39)

Today now, this program as an official name is ICPC, this program not only calls 12 countries, it spans 31 institutions, collectively they are going up to 13 cancer types. They are not all different cancer types, some do overlap, but that is fine because the dream, the vision that otherwise happen this is that ultimately when the US produce their database to me ultimate that understand cancer. You really want to make it representative of the diversity of individuals and the diversity of their cancer itself; is that combination I think they were able to better understand the disease other part of a global scale. So, what has the program done? The past talk once; so, here you go.

(Refer Slide Time: 28:22)



So, the last 12 months along, these are some of the activities that the program achieve. The very first dataset that we release to the public actually comes out of Taiwan. It was a call of standards that we did on for cancer because of the gold standard that happened in June. That are put in the public domain. We also welcome at that point 2 other or 3 institutions in two countries. Early we called human body creative university and of course, as some genome mentioned the other country that were part important was India in May of a 2018.

We also held a series of global cancer Moonshot round table sessions, basically raising awareness within each country and within their governments that helps them raise money to do the research that is very difficult to do, which is one of the good things that we are doing in today and I think we should be having a lot more these to raise more awareness and the funding. The only think I am error doing is we actually launched the training program of students; we pilot that with at this point Australia with Macquarie University.

And the other thing that we are starting to do which is quite nice as that, we are starting to take some of our laboratories and we are starting to convert them to become clear certified. So, they can take the actual test when they develop a targeted based assay and take it directly back to their two more norms. And potentially begin to further fuse together the genomic panels on with proteomic panels in influencing how best to actually treat the individual itself.

(Refer Slide Time: 29:50)

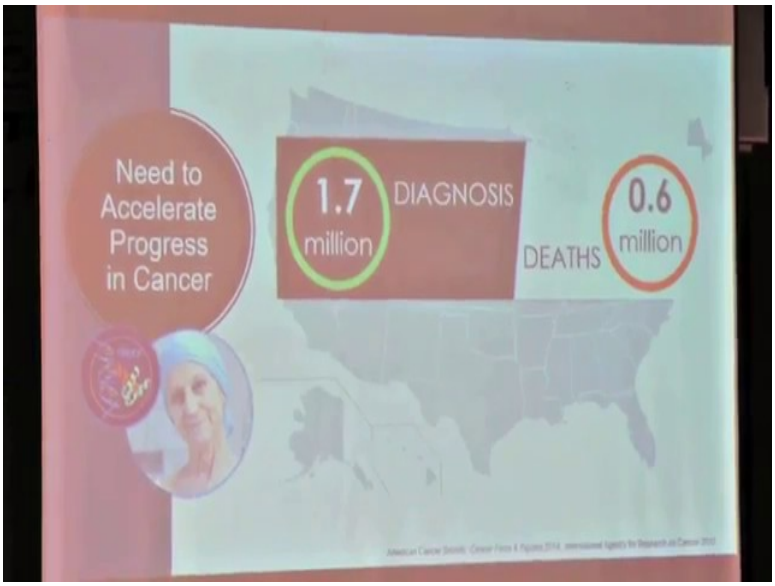
The last time we all got together, it was in the United States in Orlando Florida, just took place a couple of months ago. And as you can see it is a great family event, I have to say a

big traditional that we started. Each represented from each country, we all hold flags as a sense of pride. But again it basically shows one thing that I have learned from this is that two and a half years ago when we thought about this idea. And then two years ago when we actually launched it, everyone would say you will never entertain place in the public domain.

It is happening, we are starting to release it and there is other cancer data sets that we are released within the next 6 months once those manuscripts get accept it. There is really no barrier which is what I am learning. If you simply ask and you actually are very conscious and you are very clear about what you expect things can happen.

So, let me leave you with this final thought. So, I think a lot of promise has been made when it comes to genomics. So, I am the person to admit that because I see how that how definitely benefits patients. However, here is the reality of the statistics today.

(Refer Slide Time: 30:58)



Within the United States on average on a yearly basis just try of 2 million individuals will hear the words you never mind here. And I tell everyone that is in a research lab ghosted or training in a cancer hospital because you really understand the impact of what patients go through one early treatment and the ones that cannot survive from the disease itself. They will hear the words that you have been diagnosed with cancer.

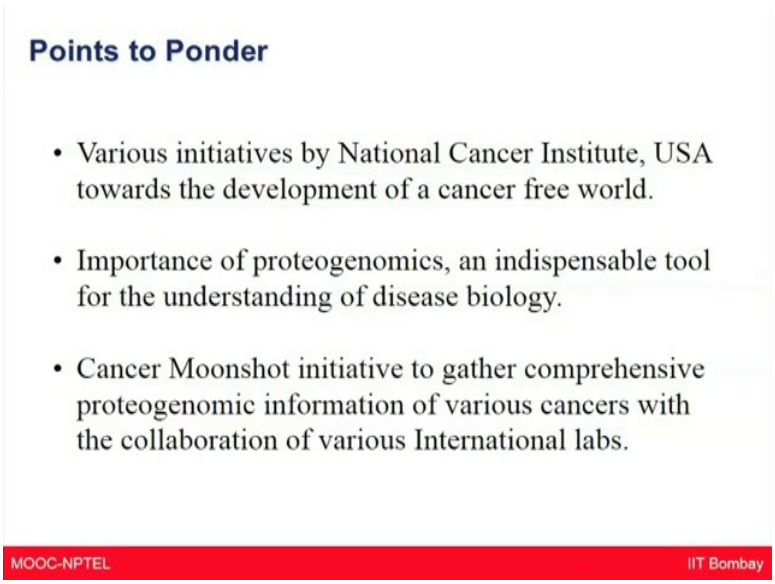
Furthermore, on average in the United States just over a half million individuals pass away from one of its many diseases, we can search this one disease it is a series of diseases that

defines cancer. But this is not an issue for the United States. This is why we develop ICPC and I think it is great that we have India on board. This is a global issue. On the global basis itself on average of a yearly basis of 14 million individuals are diagnosed with cancer and these are the ones that we are able to report.

(Refer Slide Time: 31:55)

Furthermore, just try about 8 half; 8 and half million individuals will die again from his many diseases. So, while I think a lot of progress has been made, I think there is a tremendous amount of work we still need to go forward. And quite frankly I think the work that people are doing here in India combining what we have learned from genomics infusing, it now with the measurement of proteomics in the future I think fusing it with metabolites. I think that is the key when technology become very mature there is an opportunity to combine them that is the opportunity to take because you are able to get more biology out of the disease itself. So, with that I want to say thank you very much.

(Refer Slide Time: 32:34)



In today's overview session, you are provided the knowledge of the various programs run by national cancer institute in United States. It was clear from Dr. Henry Rodriguez's lecture and discussion that genomics and proteomics are complementary, and there of course, indispensable for the understanding of disease pathobiology. You were also introduced to the importance of generating high quality data, and the various efforts the CPTAC undertook to make proteomics more reliable among the research community. The cancer moonshot program aims at collaborating with international labs together comprehensive proteogenomic information of various cancers.

India has recently joined this initiative and now we have become the 12th country to participate in ICPC or international cancer proteogenomics consortium to specifically look for breast, cervical and oral cancer. We are sure that you will be able to appreciate the importance of international collaborations, data sharing and proper quality controls, if we have to understand the disease biology and fine drug targets against cancer in the future.

The field of proteogenomics is a still emerging and every day new software and new tools are being used. It was not possible to cover all of them in this course; however, we hope that with this course we are able to lay a foundation and instilling you the enthusiasm needed to take proteogenomics research forward.

Thank you and all the best.