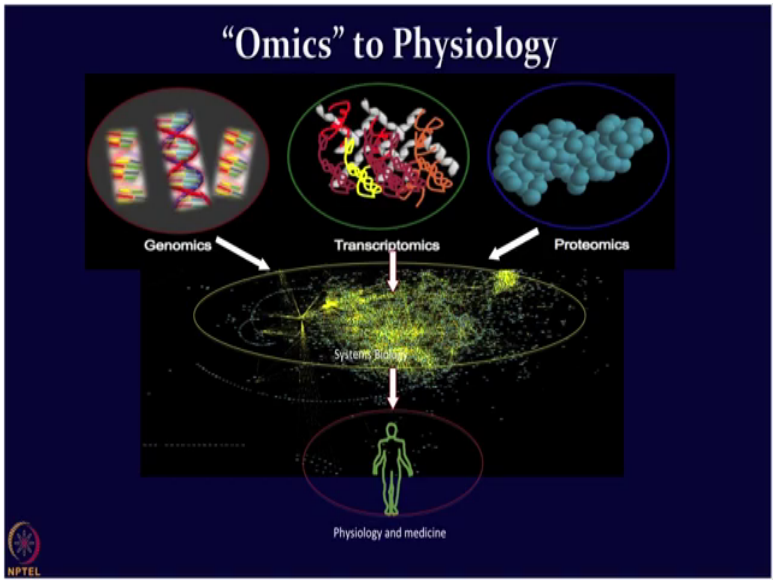
Interactomics: Basics and Applications
Prof. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay

Lecture – 60
Systems Biology and Proteogenomics- Conclusion

Hello students; so, today we are in the last lecture, we have discussed many technologies different ways of looking at various type of interaction analysis both label based and label free platforms. We also discussed about how to really start analyzing the data and visualizing the data. And finally, we have to think about how best all this omics information which we are obtaining how this could lead to the understanding of the physiology.

(Refer Slide Time: 00:46)

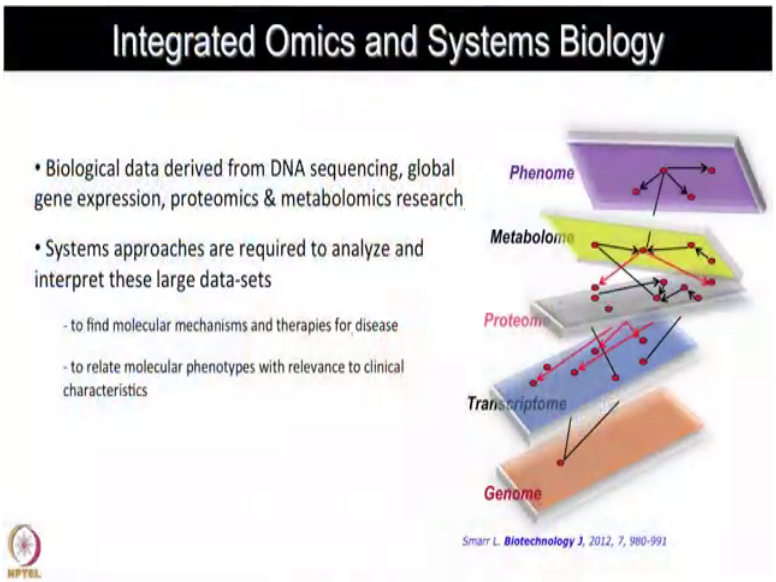


As you can see the slide that we are generating data set from variety of different level of biomolecules starting from the genome, transcriptome, proteome and then looking at this

whole information in the more systems network manner, which eventually can help us to define the physiology and probably helped the life scientists and medical scientist to really provide the meaningful information for the complex biological questions.

The complex data sets which we are obtaining from different technologies different platforms different omics levels. We have to provide a much more comprehensive view of what is the meaning of this data set. And that is where systems biology and integrated omics analysis have really started leading the hope the path of this whole field and the systems approaches are really required to analyze and interpret these large datasets.

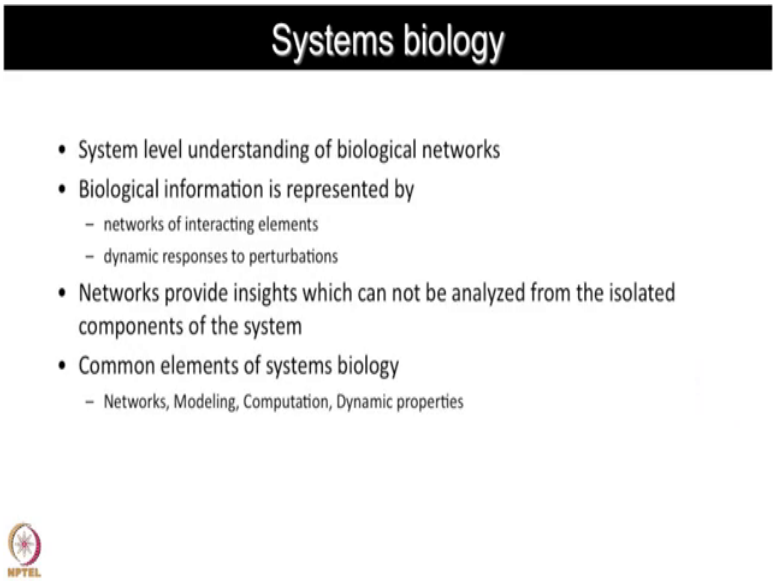
(Refer Slide Time: 01:42)



Which will eventually help us to find the mechanisms of different diseases, mechanism of different complex biological problems and various therapies for diseases. It may also help us

to relate the different phenotypes and what could be the relevance to the clinical characteristics if we are looking at more medical issues. So, what is a system biology?.

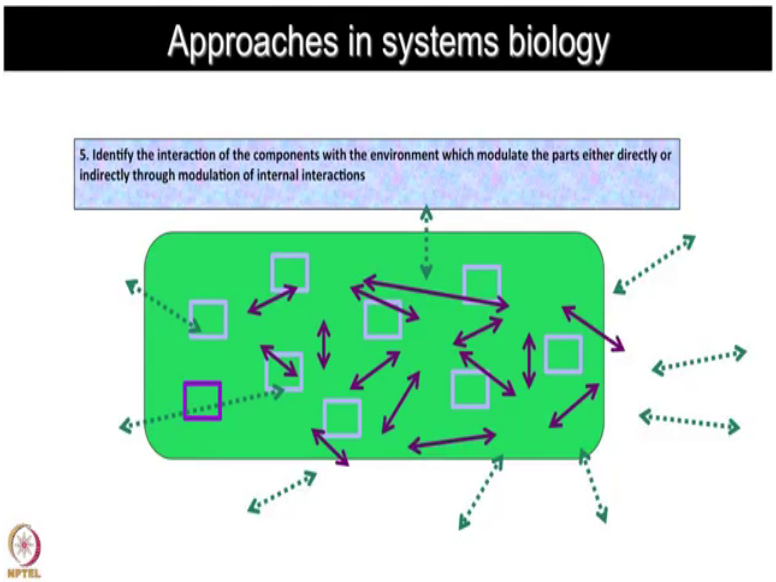
(Refer Slide Time: 02:07)



Systems biology is systems level understanding of the biological networks which is defined with the network of various interacting elements as well as the dynamic perturbations and responses to the various cues which we obtain from those systems. These networks provide the insights which cannot be analyzed just from the isolated systems alone. We cannot just look at the genome separately, transcriptome separately, proteome separately and tried to obtain the full picture.

We have to start putting the things together in a more system wise manner which could be analyzed using systems level tools. So, to the common elements of system biology, they include the networks, modelling, computation and dynamic properties.

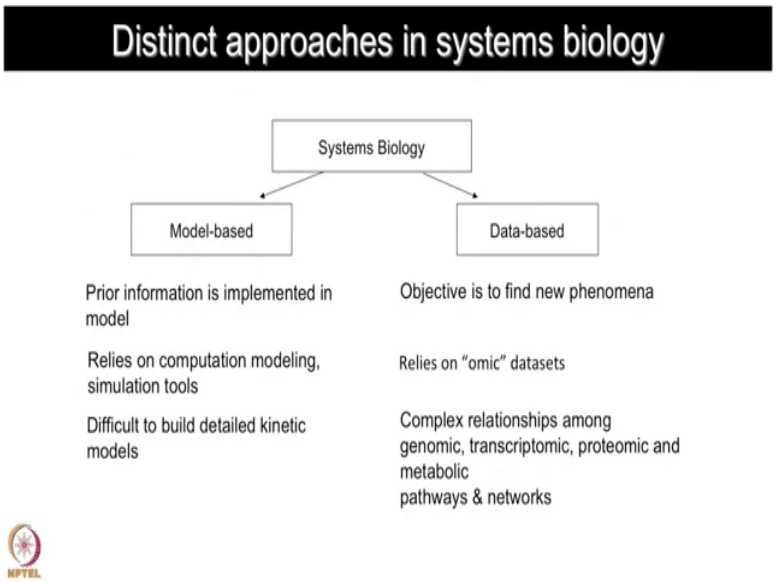
(Refer Slide Time: 02:55)



So, what are the approaches of system biology? The system is an entity which maintains it is existence through mutual interaction of it is constituent parts. You can see at the slide that system research consists of first identification of the parts then characterization of these components excluding the ones which are not the part of a system and then identifying the interaction of the components with each other.

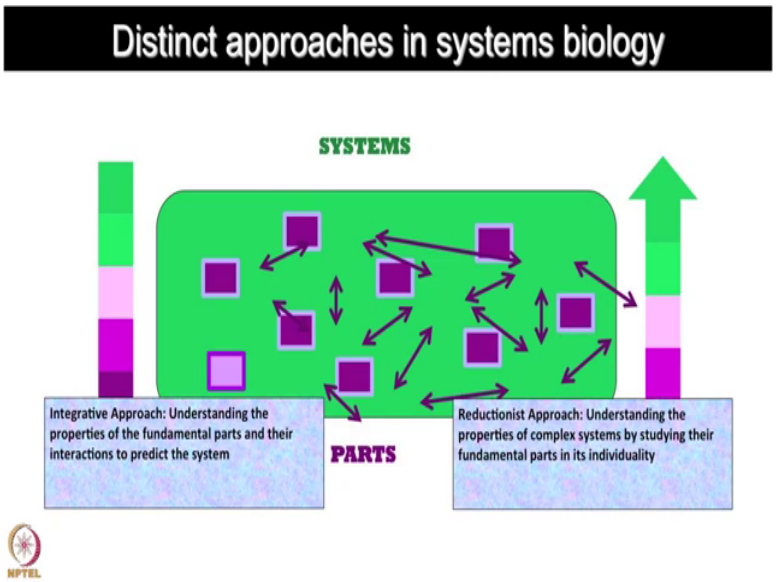
Finally, we are looking at the interaction of these components with environment which modulates the parts either directly or indirectly through modulation of these internal interactions. So, what are distinct approaches in system biology? It can be model based or data based.

(Refer Slide Time: 03:37)



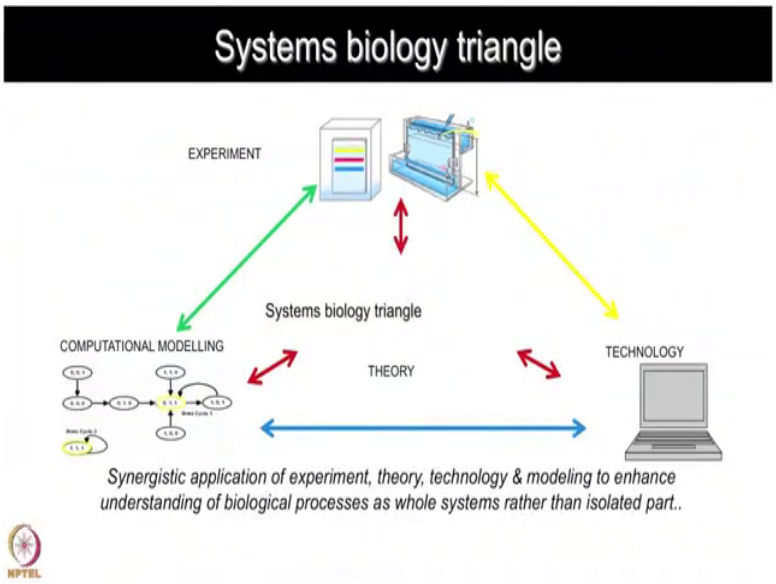
The in model based you already have the prior data information which you want to implement and build the model of course, it relies on lot of computational modeling and various type of tools for simulation, but it is very difficult to build the detailed kinetic models just based on this information. Whereas, the data based systems biology approach involves a new phenomenon to define or find using these kind of you know data sets. It relies essentially on various type of omic data sets and looks at their intricate complex relationship by looking at various components from genome, transcriptome, proteome and then try to define various type of pathways and networks.

(Refer Slide Time: 04:28)



There could be different ways of looking at the system biology approaches; one is reductions approach which is understanding the properties of complex systems by studying their fundamental parts in it is individuality or it can be integrative approach where we want to understand the properties of the fundamental parts and their interactions to predict the system.

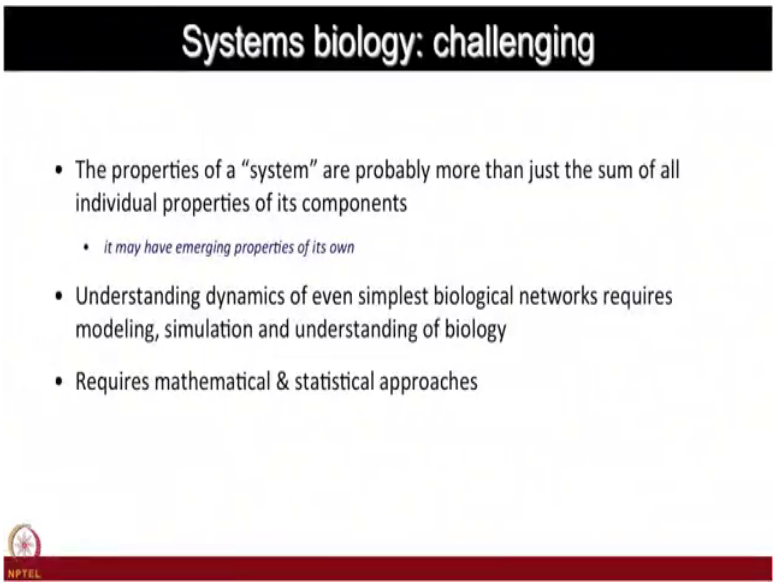
(Refer Slide Time: 04:50)



What is system biology triangle? It involves the experiment which we do in the vet lab to generate the data set, the various technology platforms and computational modeling. All this together constitute the systems biology triangle. This is essentially the synergistic application of experiment theory technology and modeling to enhance our understanding of the biological processes as the whole system rather than looking them as a individual components.

So, I must warn you that system biology field is very challenging because, you are looking at not only one type of data set in isolation, but you are trying to integrate the information.

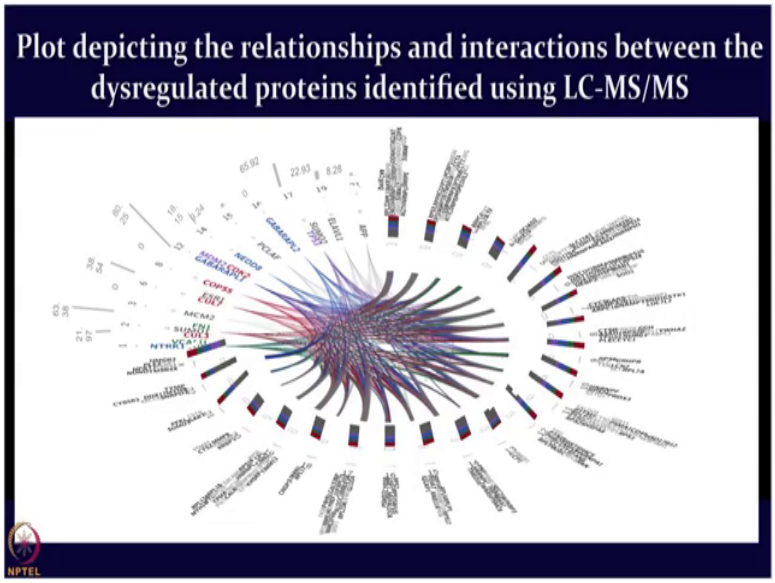
(Refer Slide Time: 05:28)



And looking at the properties of a system in it is whole which is not just you know, simply adding 1, 2, 3 and making the sum of that. But looking at much more bigger picture and trying to integrate the information in much more meaningful manner. And once you do systems analysis you might find out that system may have it is own emerging property on it is own which was not just possible by looking at each pieces separately and just by combining them together.

Understanding the dynamics of even simplest biological networks it requires lot of computational power using modeling simulation and understanding of the biological questions. It also requires mathematical modeling and different type of iterative scale approaches.

(Refer Slide Time: 06:14)



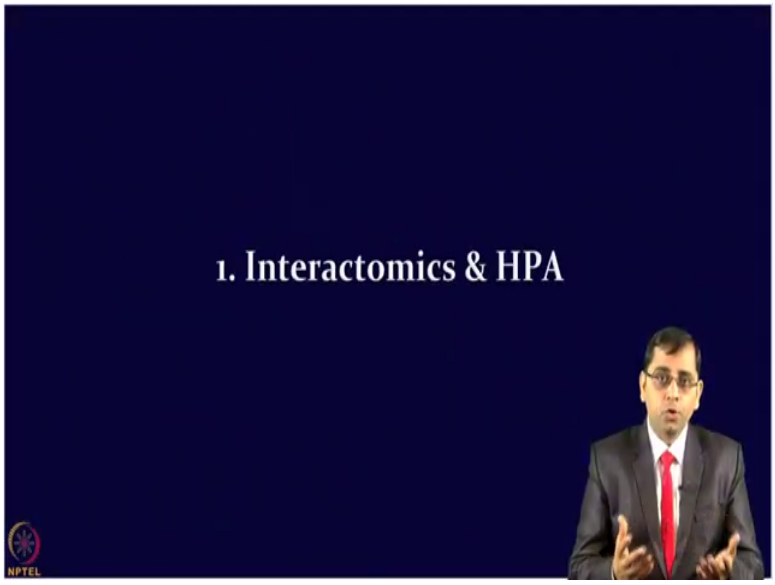
I have depicted you here the one of the omics dataset obtained from the micro arrays and mass spectrometer and then how we were trying to build the interaction networks of various proteins which were involved in a given (Refer Time: 06:26). So, by looking at this kind of information; we want to really obtain, what are the important nodes which might be governing the disease. Which was otherwise not possible just by looking at the list of the protein and what kind of proteins are changing. But rather now we are comprehensively analyzing the entire dataset and looking at which could be the major root cause of the disease and can we identify the right nodes.

If you identify those the major area then of course, one could think about certain therapies which could target those nodes and then probably those will be the root cause where one could start making more impact for the patient treatment.

(Refer Slide Time: 07:02)

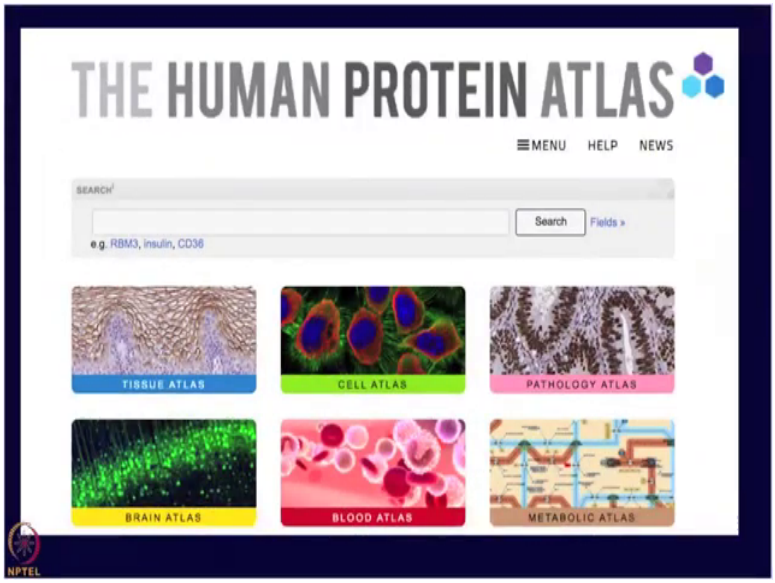
So, by next part I am going to now talk to you about some of the major revolutions which are happening in front of us in the field of omics. I am going to give you some very brief examples and also going to give the perspective directly from the experts, the scientists who are working on this field and who have actually laid the foundation for the areas.

(Refer Slide Time: 07:23)



So, first in the field of interactomics; the big project has come forward which is the human protein atlas.

(Refer Slide Time: 07:30)



The scientist Mathias William and his big team, they are all working towards generating antibodies for each protein and looking at that how the proteins are localized, how proteins interact and they are providing a lot of interesting biological insight looking at the proteins by different type of experiments which are based on the antibodies. So, let me have a perspective of Dr. Mathias William and Dr. Emma to talk to you about what are the goals of these big project of human protein atlas and what are the major accomplishments so far.

So, we have during the last 10 years as part of the human protein atlas project. We have been making at cell atlas a sub cellular map of the human proteome using an antibody based approach. So, we generate high resolution present images and we have been generating hundreds of thousands of such images and then we have been classifying the patterns and sorting it into different categories.

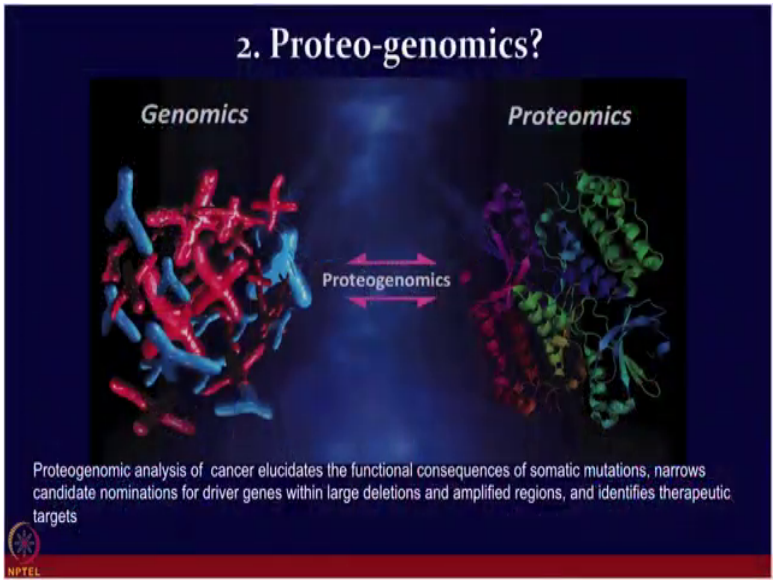
So, we could finally, provide the first map of the first sub cellular map of the human proteome and see which proteins are in the mitochondria which proteins are on the plasma membrane and so on. And the most interesting findings we made here is that as much as half of all proteins localized to several places in the cell and we also see a lot of single cell variations. So, the multi localizing proteins as we call them it is very interesting from a biological perspective.

We do not know the biological consequences of this it might be. So, that these proteins have context specific functions or even (Refer Time: 09:08) light in different parts of the cell, but we cannot really tell we only we can only observe make the observation that they are in multiple places. Then, someone else would have to do or would have to do in depth biological studies look at splicing isoforms or post translational modifications to see if it is how this multi localization is achieved and also what are the biological consequences.

My opinion is that antibodies should be viewed upon small chemical reagents, they have on and off target binding. So, you have to validate your results in any assay where you use them. So, we fail more than half of the antibodies that we use. So, we put a lot of effort into validation of our data and we have validation scores that kind of denotes the reliability of the results. So, that is a big part of our job.

Next I will like to talk to you about the field of proteogenomics. While lot of things we talked about different proteomic technologies.

(Refer Slide Time: 10:06)



But you have also seen that how NGS is contributing resolution sequencing and genomic field is contributing, immensely towards our understanding of the biology. So, how to integrate this information? Can we start integrating the most crucial molecules say genome and proteome together and that has actually led to the field of proteogenomics which is essentially going to look at you know a lot of information which was not possible, otherwise by looking at the molecules alone.

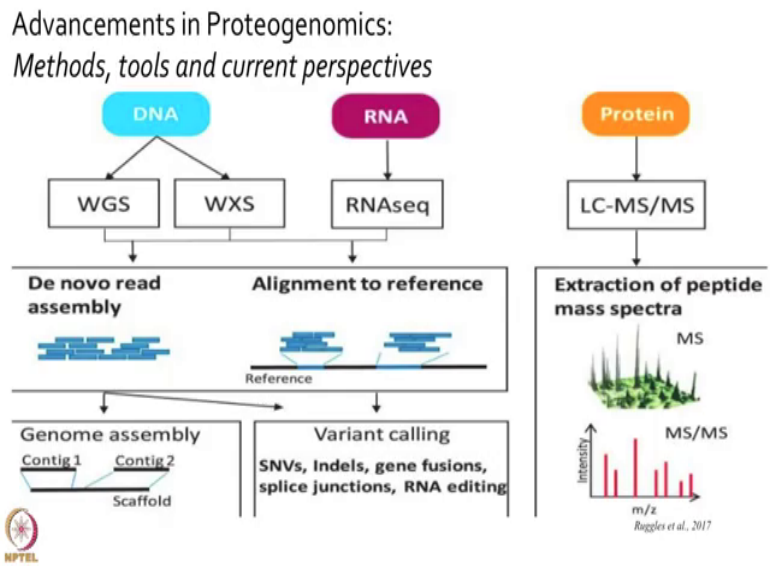
Probably with this cartoon you will appreciate that how the proteo genomic analysis of cancer and other complex diseases could elucidate the functional consequence of somatic mutations narrow the candidate nomination for the driver genes which may have the large deletions or amplified regions and probably identify the therapeutic targets.

(Refer Slide Time: 10:52)

The initial cancer institute and national institute of health in the USA; they have really made huge contribution in this area of proteogenomic investigations. There are series of interesting paper which I have shown on the screen. They have showed the utility of integrating these technologies and how one could actually understand the very complex concepts in much more efficient manner. Especially the studies on the collateral cancer, breast cancer and ovarian cancer have really given the path that how to integrate this information and get much more meaningful insight for the same complex disease which was otherwise not possible.

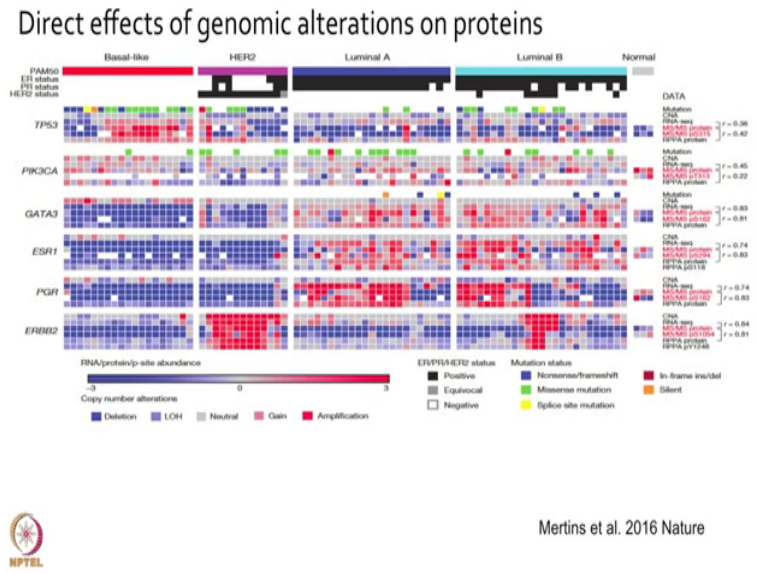
So, in the field of proteogenomics, which actually we offer another course on this is a much complex subject and requires much more discussion, but to give you a flavor it involves from the same sample and it is more used in the case of cancer and other complex diseases.

(Refer Slide Time: 11:49)



From the same patient sample can we look at DNA and look at their whole genome sequence and whole exome sequence, then look at the RNA analysis RNA sequencing in transcriptome analysis and protein and analyzed using LC, MS, MS and then start aligning the data I start analyzing the data big data to obtain the proteogenomic information which has given those paper which I just showed you in the previous slide.

(Refer Slide Time: 12:18)



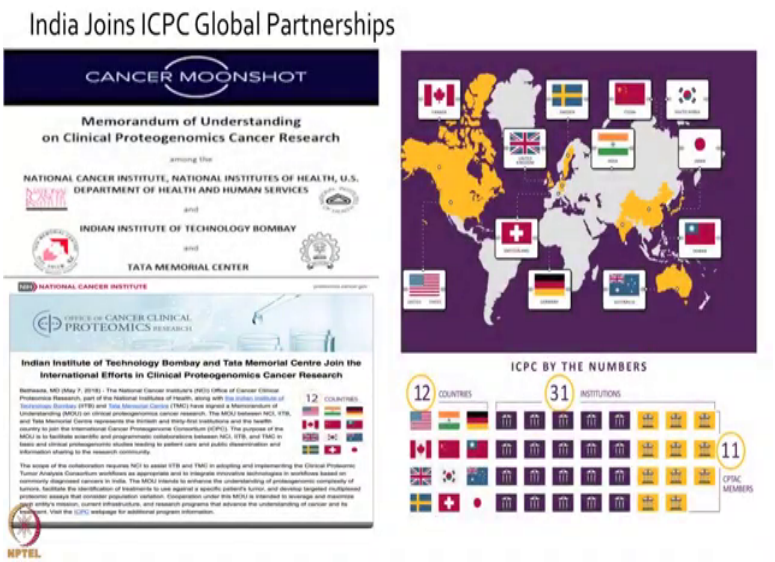
There is a many beautiful studies which have shown the impact of these studies and they have also shown that you know how if you build the layer wise information as you can see in one of this nature of study shown in the breast cancer, that if you are looking at different subtypes like basal like, her two Luminal A, Luminal B; these cancer type if you are only looking at one layer of information only at the gene level or even only the protein level that may not help you to obtain the entire picture.

But even some time information of the phosphorylation level or PTM level is much more powerful than just looking at all of this other molecules. So, you have to you know get an idea that how this entire information can be put together and can give us much more meaningful insight. So, based on this idea that how to utilize these omics field omics technology the ex US vice president Joe Biden.

(Refer Slide Time: 13:04)

He started the project cancer moonshot. Idea was can we accelerate our pace of doing cancer research and whatever we are able to achieve in next 20 years can we achieve in 5 years. So, in this manner, scientist really got motivated and started coming together in the field of proteogenomics which led to the cancer moonshot USA program. The next intention was cancer has no boundaries. Can we expand this project to the international community and they form international cancer proteogenome consortium to have more association and collaboration from other countries.

(Refer Slide Time: 13:37)

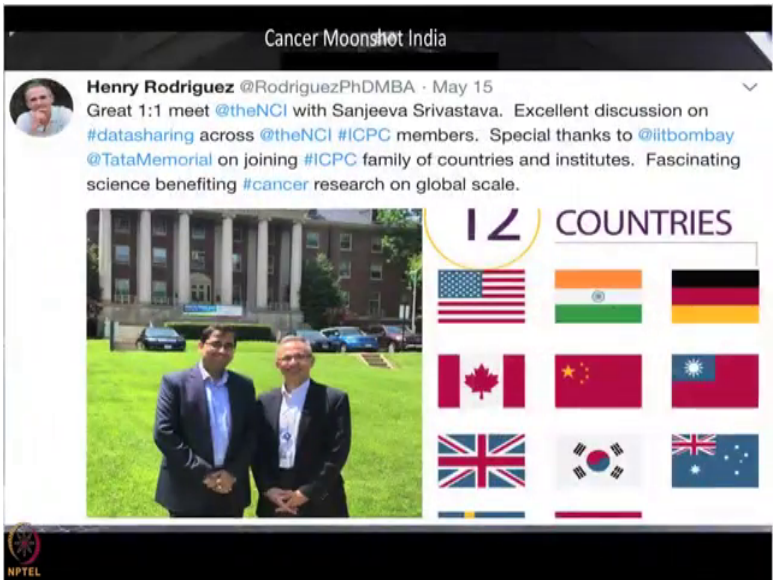


We are very fortunate and proud to be part of this program of ICPC where India joined the 12 country consortium and we are also going to look at the proteo genome analysis of different complex cancer of India and then share the data and analyze the data with different countries and try to obtain more meaningful information for these difficult to treat cancers.

(Refer Slide Time: 13:59)

An image shown here for one of the ICPC retreat where different countries are participating in this activity.

(Refer Slide Time: 14:06)



So, within the picture, I have shown Dr. Henry Rodriguez, he is one of the director of NIH, NCI who has really led the path for the ICPC program. And I am going to show you a video clip of sharing his perspective about the ICPC program and cancer moonshot as well as what are the guidelines for doing research in this area and the way forward his vision.

Speaker 3; so, what I could speak to is to more from the national cancer institute which is the effort that I leave. I think the part that is very exciting I have watched now the evolution as I would call it and the maturity of the science of proteomics.

So, 10 years ago when we started this effort at the cancer institute referred to as now the clinical proteomic tumor analysis consortium CP TAC those were an acknowledgement out of NCI while we saw the potential for proteo genomics. We knew proteomics would be a very complementary field, but at that time we felt that technology was being developed in this case

mass spectrometry yet while it is extremely a very powerful technique there were items that we needed to address in terms of the rigor and reproducibility if it were to one day be a complementary technology to the genomics landscape.

So, 10 years ago while at cancer genome atlas was the very first one large scale biology that moved forward in cataloguing all the perturbations inside cancerous cells from a genomic perspective. The proteomics field which is already take a little different approach. We went after the analytical rigor of those technologies take about 5 years. But once we got that finally, on hand the question became what would we do next.

There was the opportunity that we seized we went after those same tumor types that should not that the genomics community went after the NTCGA and we applied those proteomic base technologies mass spectrometry onto them. When we saw the beauty of the additional biology is going to be generated from the combination of these two disciplines that is when we saw a lot more potential.

So, today this program now in it is next iteration in ways it is it will be expanding we will go after additional cancer types from the first program. So, the first program we went after colorectal breast and ovarian cancer. Now we will be cataloging approximately additional 5 cancer types, but the most exciting part that I would have to say for the first time, we are going to be teaming up a proteomics laboratory expertise with an ongoing NCI sponsored clinical trial.

And these clinical trials are typically genomically driven, but we recognize that genomics is still advanced, but there is an opportunity to fill in the void of the biology of trying to understand why patients either respond well or do not respond well. Based on a genomic driven information to the treatment that they were just administered to. I would say the best way that I would look at it is what is driving precision medicine today.

In oncology which is my specialty I think clearly genomics is making tremendous strides there are exquisite examples today where we are able to identify and be in a very great position and how to treat a patient. Simply rely on genomics; however, you kind of flip your

coin there are still many instances in oncology we do not fully understand the biology and how these individuals are responding to the treatments that is where I see now proteomics filling a very critical piece of that information.

So, what we are trying to do at the NCI through our initiatives is begin to converge these disciplines. And quite frankly partly, I think the reason has been is the evolution of technology. Genomics technology has matured and is still maturing at a faster rate, but as proteomics technologies are also maturing and also we could go after smaller samples it is just a matter of time that these disciplines do begin to converge.

So, that is why I see these two areas being very complementary to one another. Specifically an area precision based medicine for oncology as opposed to the common problem of mass effect I would say what is needed. Because I really do not see the most problems I just see them as the evolution of technology. My one wish I think which is a natural wish that you have of all technologies a common ones you want automation, you want it to be very simplistic for the average user to use. You also want that the system as technologies of all to potentially using less sample material up front.

But keep in mind I think this question is one that people tend to ask on is well how much material do you need to do the analysis that you wish to go after. The question that people should be asking is what do you what is the question you are trying to ask with biology. That question will dictate what you want the technology to do.

And in most instances you find out when people really begin to ask the question they would like to see if the experiment technologies themselves are quite mature from what I have seen from my perspective proteomics the way we know it today would not exist without mass spectrometry pure and simple genomics gave us a blueprint what people did with in the specialty is there was an acknowledgment here is a mass spectrometer.

Mass spectrometry has been used in clinics for over 30 years for different purposes. It is mainly used for small molecule screening, but then a bit, but then there was a brilliant idea that came into play. Can you take a big protein break it apart into small pieces into these

peptides and informatically, just like an eloquent puzzle stitch it back up into a protein without mass spectrometry proteomics as we know it today simply would not exist.

So, the cancer moonshot I would say is one of the most incredible things that is been developed off the united states over the past. Now what approximately 1 year. So, I will see one of the main drivers here is the 47 vice president of the united states Joe Biden it is been a huge honor to be part of that effort two programs that I could specifically speak to myself on one of them we actually did which is a partnership between the to the national institutes of health specifically NCI, then we partner with the department of defense.

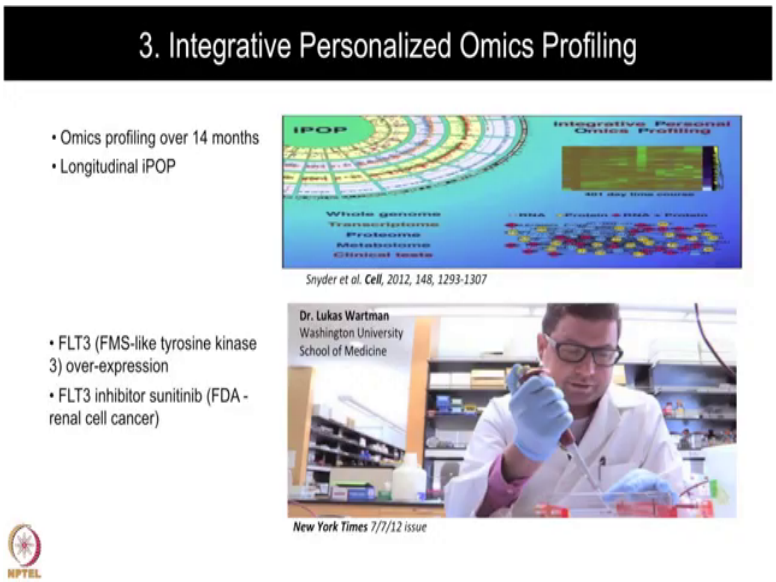
We also part of and we also partner with the veterans administration. This program is referred to now as Apollo this effort along with another program that we have now launched under that cancer moonshot umbrella is an international program. This one I was referred to as the international cancer porteo genome consortium or ICPC.

That program now which is an incredible effort in itself. Actually, now involves 27 institutes spans 11 countries. What is nice about it fills one of the key promises that our former vice president wanted. And that is the data that we generate in this international effort will be placed in the public domain. But the part that is nice about these two initiatives is that each one of them.

ICPC and also the Apollo, there will also be given up the data in the public. But what brings them together it is the foundation of what CP TAC actually did and that is a recognition converged genomics with proteomics with the hope that only you will identify a new biology that is in the Apollo case. They really want to push it much more faster in the translational space potentially in the near term future begin to apply it towards patient care.

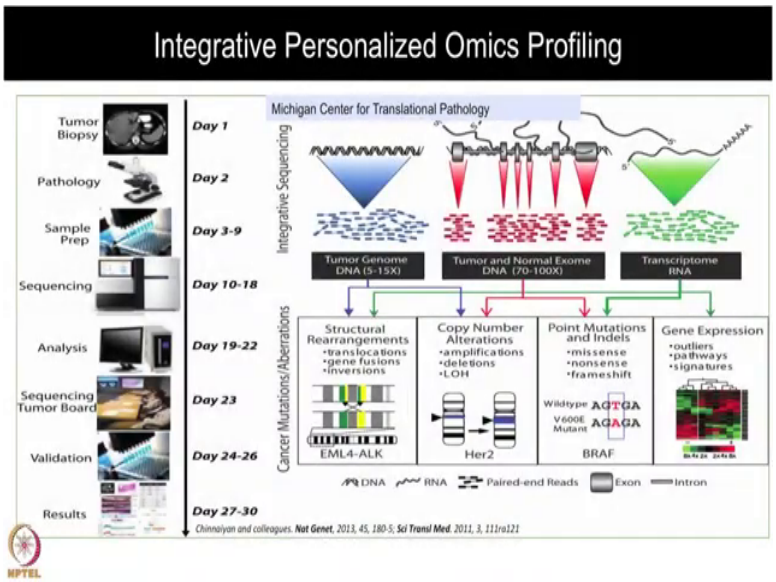
Another interesting development happening in the field is integrative personalized omics profiling. Essentially this path was led by Dr. Mike Schneider from Stanford University who took his own samples and analyze longitudinal manner using various multi omic technologies which really led.

(Refer Slide Time: 21:57)



First time idea that you know he might be prone for diabetes which was detected clinically very late, but looking at the biomolecule, they could sense this much ahead of time. And this really constitutes that how looking at the multi omics analysis from the same individual can provide more meaningful information. And likewise, there are many success stories many different case studies are now coming forward which are showing the utility of looking at personalized omics profiling.

(Refer Slide Time: 22:35)



There are also groups in Michigan who are looking at Michigan center for translational pathology Dr. Rocha Naen and team. They are doing from the same patient looking at the genome data, their proteome data, their metabolome data, how one could start finding the right cure for the given tumor type which is otherwise not possible. And they are able to provide you know at a much comprehensive level this kind of analysis and directly reaching to the patient care.

Now Dr. Mike Schneider and his team has launched a major program on human personalized omics profiling or H PoP and idea is to provide much deeper understanding of the genetics of the human body and provides the new clues of different diseases which was otherwise not possible. Let us hear Dr. Mike Schneider about his program and vision for the H PoP project.

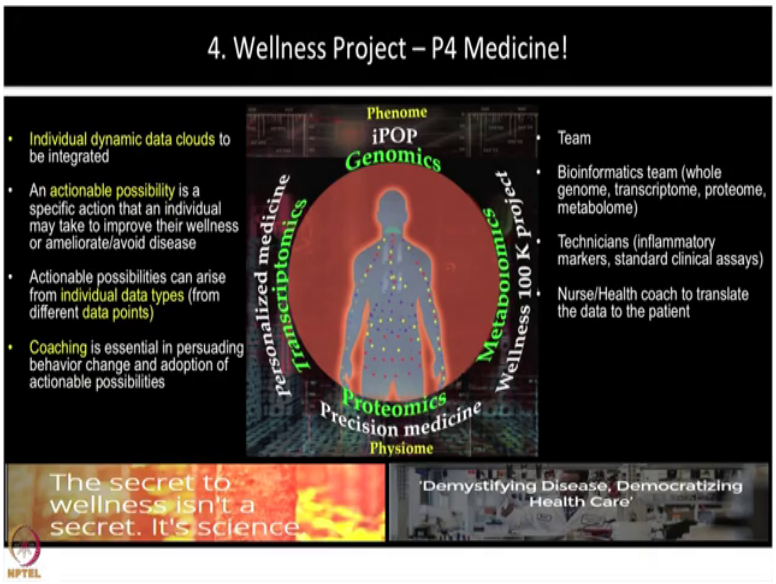
Speaker 4: what is the overall goal of the H PoP project H PoP project is really trying to understand what it means to be healthy and how do people differ all around the planet. So, what we are doing is we are actually taking peoples blood and urine and also their stool samples. And what we are doing is? Doing deep omics profiling the proteome, the metabolomes, the transcriptome to make as many measurements as possible to see at a molecular detail how people differ from one another.

So, we are going to see this, we are going to sample people from all around the globe at these human proteome meetings and then we are going to see what their molecular diversity looks like how two different ethnic groups and different locations have an impact on peoples molecular composition. Do you think omics will be used in clinical self situations? I think at some level absolutely that is to say we will be making thousands of not more measurements in the future.

Right now when you go to a doctor they make about 15 measurements on you and in the future I think we will be making hundreds of not thousands using various omics methods. Especially proteomics and metabolomics will be very very powerful for looking peoples molecular state which I think will say a lot about their health and also what happens when they get sick.

Speaker 1: let me now take you to the next project which is wellness project. We have been talking mainly about how to compare the healthy individuals with disease.

(Refer Slide Time: 24:45)



But can we look at the wellness or the health status of healthy individual or seemingly healthy individual in the longitudinal manner. So, can we look at their entire omics profile at different time interval from 0 time to 6 months to 1 year, 2 years and start looking at are the changes happening in dynamic manner inside the body which might give us the clues for what disease the individual might be having or whether the individual is still healthy.

So, the individual dynamic data clouds has to be integrated and that field is known as P4 medicine, which is led by Dr. Leroy Hood from institute for Systems Biology in US in Seattle. This big project of wellness was also published 2 years ago in nature biotechnology has shown the agility. And many similar projects have not started in different countries which are looking at the omics data set to build the wellness. Let us hear the views of the Dr. Leroy Hood on the wellness project and P4 medicine.

Speaker 5: I am very excited by the possibility Indian, India would consider a scientific wellness program starting with a pilot project and it is own population. Scientific wellness I think is the keystone for the revolution in medicine because through a longitudinal analysis of populations using what I term personal dense dynamic data clouds that is the analysis of genomics of proteomics of metabolites of clinical chemistrys of digital assessments of activity and sleep and things like this.

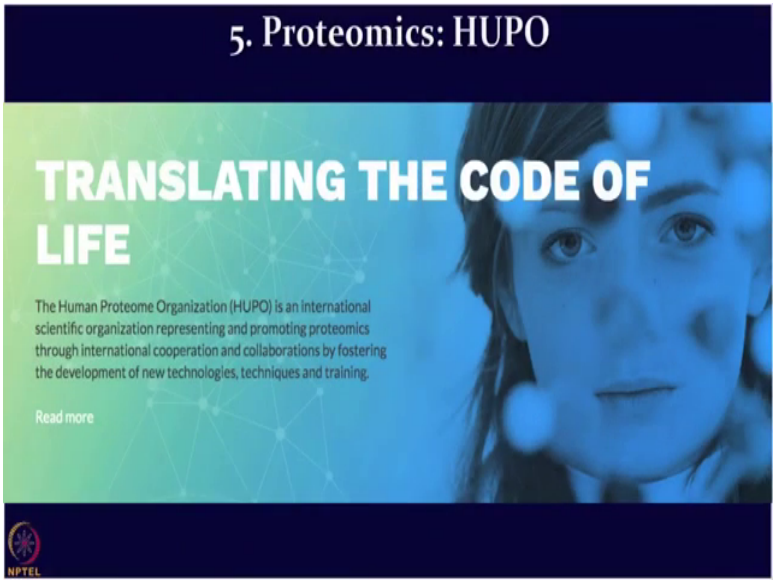
All of these together when analyzed properly can lead to actionable possibilities to improve wellness and or to avoid disease. But I think the really transformational possibility scientific wellness presents is that if you create a population of people that follow the scientific wellness program in a longitudinal matter.

And if that population begins to right reach a certain critical level of thousands in and we begin to see wellness to disease transitions. That you can use the dense dynamic data cloud to get biomarkers to mark say the earliest transition for diabetes. And then you can begin to think about the systems technologies and systems driven strategies that can be used to identify the disease which has networks that will reveal candidates for creating drugs to reverse these diseases that are at a very earliest age.

And if for example, over a 20 year period, we can reverse all major diseases before they ever really manifest themselves as a disease and become irreversible. We will save the health care system an enormous amount of money. In the US, for example, 86 percent of our enormous healthcare budget goes to chronic diseases suppose in 20 years.

There very few chronic diseases left we have transformed a healthcare system that thinks only about disease to a health care system entirely focused on optimizing wellness for their population. This is the opportunity that India has and I would be delighted to help you in that endeavor if you decide to make the commitment.

(Refer Slide Time: 28:46)



So, finally, coming back to the proteomics. And the field of proteomics is now very well led with the efforts of human proteome organization which tagline is translating the code of life. Idea is many scientists working in the field of proteomics and integrated omics how to come together they start sharing the data, you start sharing the ideas, exchange their various information build collaboration and come forward for the big projects.

And human proteome project is one of the ambitious project which is still happening which is looking at the chromosome centric human proteome project or disease and biology centric human proteome project and now there are new pilllars added for the pathology and data driven human proteome projects. So, let us hear the detail of HPP project with ex president of (Refer Time: 29:36) Dr. Mark Baker.

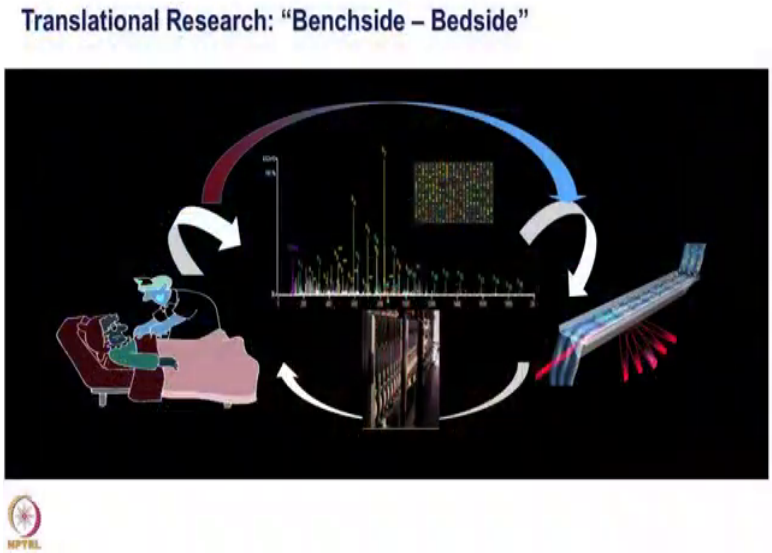
Speaker 6: the human proteome project has two major goals. The first is to map the expression of the human genome in terms of which proteomes are made and which post translational modifications are on those proteomes that affect it is activity. And the second is to springboard discovery in new drugs, discovery around biomarkers for human disease and an understanding of human health and wellness are one of the major activities of the human proteome project has been to determine how many of those proteomes. We have observed in human health and human disease.

So, far we have taken an approach that looks at about 85 percent of the total proteome content of the human proteome, but out of the 15 percent that is missing about 7 percent of them we have some evidence for them that is either mass spectrometry evidence or evidence that comes from other types of scientific experiments, biochemistry, physiology, pharmacology, genetics. But there are still about a 1000 proteomes in the human proteome that we have absolutely no scientific evidence for their existence whatsoever.

Proteomics is not a word that everyone in the street knows if I was to explain what proteomics was it is a effectively the proteome complement or the machines that are made from the human genome from the coding instructions into something real that makes a human body work. So, the proteomes are those things that are important in the machinery of life in the reproduction of life and in the development of disease that is why it is important for us to understand the proteomes as well as the genes. In order for us to understand human biology and how that goes away in human disease.

Additionally, it will be interesting to also listen the last 3 Yupo presidents; Mark Baker, Mike Schneider and Stephen Pennington and hear from their perspective what proteomics and Yupo has achieved so far. And what are their vision for the future guidelines for doing research in this area. I hope this will be interesting conversation for you to know the perspective of these leaders who are leading the field of proteomics.

(Refer Slide Time: 32:09)



So, finally, I will stop here to show you the slide which is a motivation for many people working in this field of omics and life sciences where lot of you know the actual problem. We want to find the clues and the mechanism and the leads understanding them at the omics level at the molecular level and then want to translate our bench side findings to the bedside for the actual patient care or for the actual welfare system to really make impact of that. And that is where all of this understanding all of this technologies are really going to help us to achieve that major goal.

I hope you this whole course the different type of life sessions, different demonstrations were helpful for you to appreciate the power of these omic technologies. It has definitely given you more confidence that you can start doing many of these experiments and projects even in this small colleges and other universities, where you may not have availability of these big gadget and platforms with the access of so much data available being shared by the entire

community. I think you can start looking at and appreciate the data in a very different manner.

I hope the whole course has given you more understanding much more new knowledge about the field of interactomics more you know in detail as well as the broad understanding of the field of proteomics and other omics sciences also you are now more confident to start doing some experiment and doing project in this field.

We will be happy to have more interaction with you which we will be doing further. And I hope that you know some of these understanding and knowledge is going to help you not only in your research, but also in your future projects and for your carreer. Thank you very much for attending this lecture and attending this course and wish you all the best for your future endeavors.

Thank you.