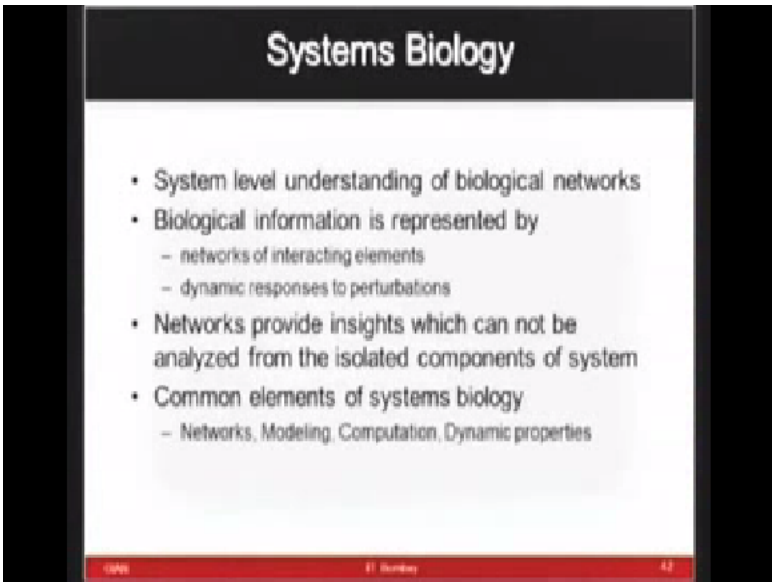
Introduction to Proteomics
Dr. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology - Bombay

Lecture – 38
OMICS and translational research

Let me now switch gears and give you the kind of a bigger vision what is happening throughout the field, not only what is happening inside the room looking into the quantitative proteomics, okay. So that is just the one part of the bigger picture what is we are studying here but if you really want to obtain meaningful information, you have to be much more broad.

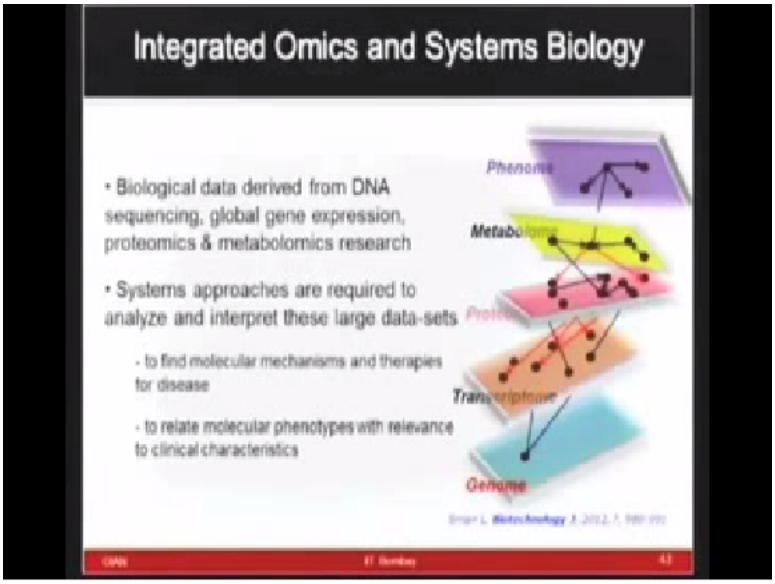
And therefore as I mentioned, my lab uses variety of platform, not only the mass spec-based platform but even within proteomic, there are variety of platforms which one need to be very open to to use and depending on what question you want to ask, you can utilise that right platforms.

(Refer Slide Time: 00:57)



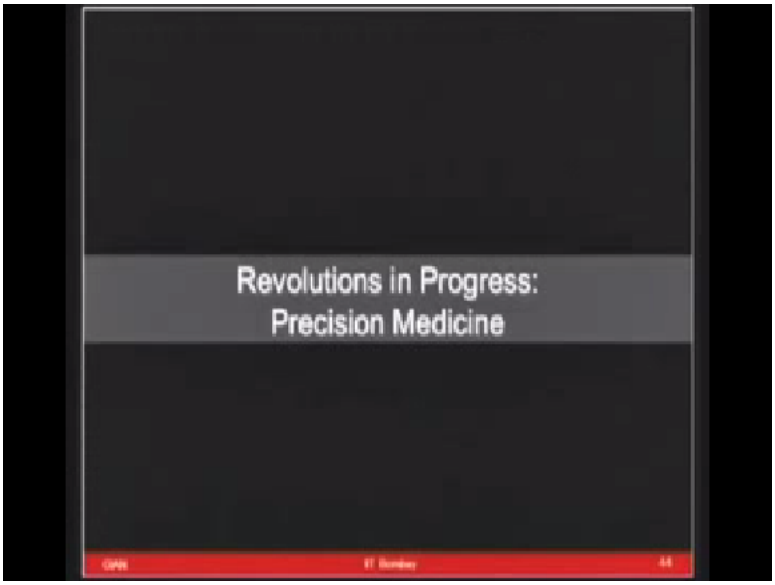
Right now just mentioning to you like the systems network information which one could obtain from variety of ways of networks, modelling, computational-based compression information and these systems network can only come if you have the access to the large OMIC kind of dataset at the variety of levels.

(Refer Slide Time: 01:18)



Especially if you think about different approaches which I talked in the very first lecture as well that if you have information from the genome to transcriptome to proteome, metabolome and phenome, then probably you could build that information at the systems level and to glean into the actual biological problem, one need to do lot of computational analysis, modelling, analysis, building those network and then you can find the right answer to the problems.

(Refer Slide Time: 01:39)



So what I will do now, I will kind of just give you some very intriguing thoughts about what is happening elsewhere in the world and the kind of revolutions in progress and I think we are very fortunate, we are in an area or in an era where you live where all kind of new things are still in the process, okay. It is not that you have already learnt that, you can do a cloning and you can do

how to run the cloning from the sample book, right. So it is not that.

You have to really still develop new assay; you have to still develop new technologies to find the answer. So you are at a part when you are the part of the revolution in progress and there are couple of things which I thought to mention to you here, one is precision medicine, okay. Precision medicine, is how accurate your targets can be to define the right disease stage and therefore in US, they have very recently launched a big initiative which Barack Obama kind of made announcement. I will just show you this highlight here.

This is going to be completely unprecedented. It is going to affect medicine in ways we have never seen before. You can match a blood transfusion to a blood type. What if imagine a cancer to a genetic code was just as easy, just as standard and that is the promise of precision medicine, delivering the right treatments at the right time, every time, to the right person.

Precision medicine enables the ultimate delivery of personalised medicine which means you come to your doctor, you have your genetic information and the doctor can use that genetic information to specifically design treatments for you, but obviously, it will be thousands of individuals that have your disease and your gene and so it is personal for you but it is also treating large numbers of people in the same situation.

We can look at individual tumors, we can even look at individual cells within tumors and understand how their system has gone array. It is possible to analyze this very complex data using state of the art computer algorithms to be able to see the interaction proteins on a cancer cell and to design drugs that will inhibit or treat them. A really good example of how this works is with a drug called Gleevec and this is a drug for a certain type of leukemia.

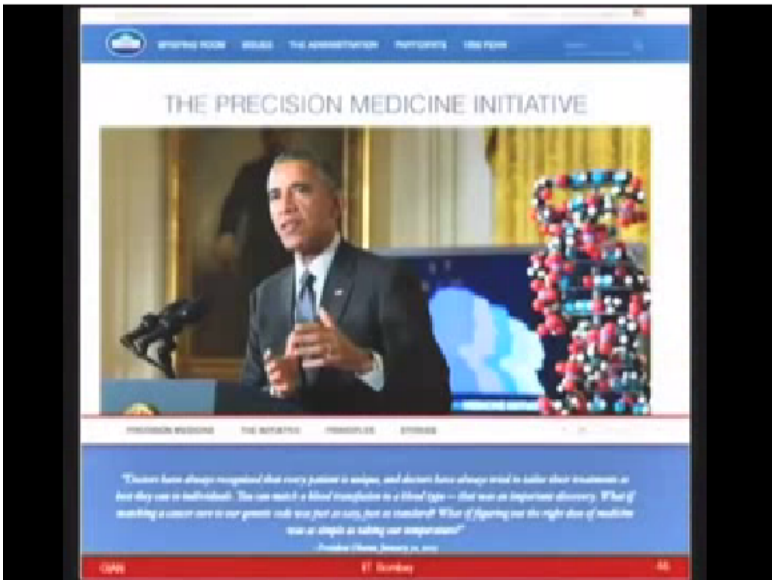
This drug had a miraculous effect, I mean it basically cured the disease. However, the individuals who receive this 6 months or year later, was back. Then a second drug was developed, we had the tools to go back, look at the sequence of that reverted protein and know where the change was and then designed a second drug to address them and now the combination of those 2 drugs that are used in sequence basically has cured one form of this disease.

And so I think that is what is going to happen, there will be more and more examples of cases where we can join genetics biology and medicine. Alright, so not only proteomics is the ultimate but of course there are various complementary technologies which has made significant progress and genome is of course has been leading in that way, being all the genome sequence in around 2002 or so.

And then availability of next generation sequencing platforms now which has made it very robust platform now and the cost has also got drastically reduced to sequence each genome in less than $1000. So based on that precision medicine currently is more aiming towards looking at the genetic mutation and individual variation at the genome level and US government Barack Obama recently announced this particular precision medicine area.

Whereas of course looking into the facets of the precision medicine using genomics, precision proteomics is the next way when proteomics field is progressing and of course lot of things which you are learning in target proteomics is catering towards that goal.

(Refer Slide Time: 05:50)

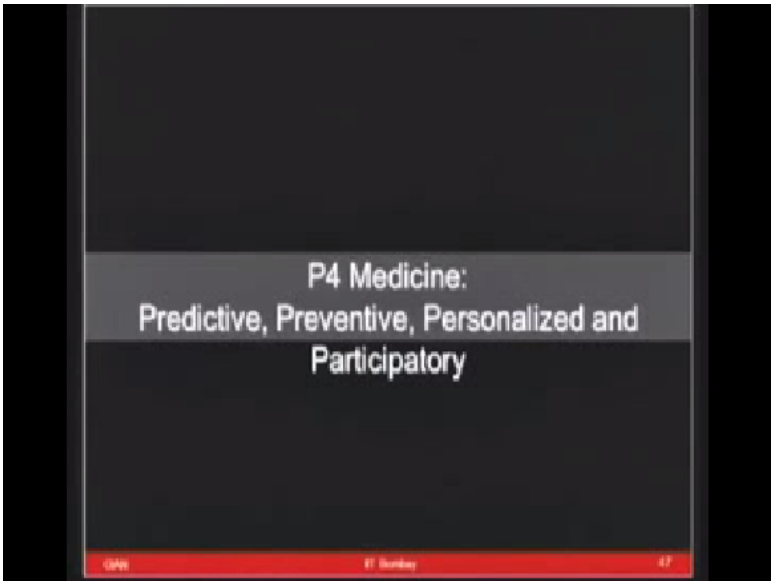


So this is what precision medicine initiative Barack Obama mentioned that because each individual, each patient is unique, doctors are always try to tailor their treatments as best as they can do as an individual. You can match a blood transfusion to a blood type that was a very

important discovery but what if matching a cancer cure to our genetic code was just as easy just as a standard.

What if figuring out the right dose of medicine was as simple as taking our temperature. So far these problems, these have been very complex. We cannot measure them as easily the way you do any blood test or you do your temperature measurements but if the individual patient therapy lean precision medicine be possible which looks like it is going to be reality now. Things will become very different.

(Refer Slide Time: 06:37)



Another initiative is happening is P4 medicine, looking at the systems-level information again. The way the field is progressing towards the predictive, preventive, personalised and participatory and the major lead for this part of project is coming from Institute for Systems Biology, ISB at Washington and Dr. Leroy Hood, he is the pioneer in that area. Let us listen from him the emergence of P4.

Over the past 12 years, there has been a profound revolution in biology, that is we have started making systems approaches to the complexity of biology, approaches that are holistic and global in their nature and the application of this system is linking to medicine which started perhaps 10 years ago or so, has created both what I call systems medicine and the application of the systems medicine to patients, something that I call the P4 medicine revolution.

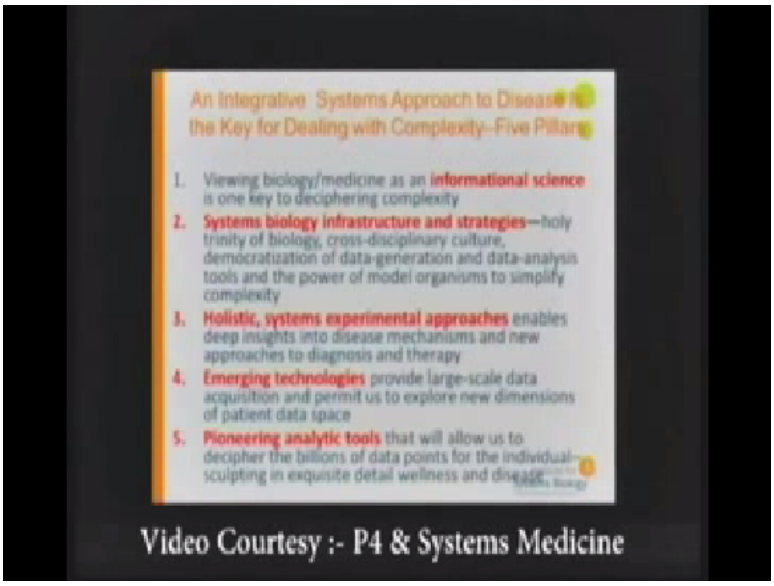
And a really important point that I would like to make is the P4 revolution personalised medicine really on steroids comes not only from the convergence of patient-driven social networks and the convergence of big data, lot of which you have heard about here but it comes from the convergence of systems medicine. I think it is really important to point out that in many of the national publications are the creative destruction of medicine by Eric Topol.

He beautifully outlined the network and the big data aspects of the new medicine, he did not talk about systems medicine. So what I would like to do is give you a real sense of what systems medicine is all about and how it translationed the patients in terms of P4 medicine is really going to be revolutionary. My view in the world is that each individual patient in 10 years will be surrounded by a virtual data cloud at billions of data points of enormously heterogeneous types ranging from molecular, all the way up to social networks.

And the reason we need this enormous amount of data is the incredible complexity of biology and by inference disease the complexity that arises from the normal chaotic process of Darwinian evolution. The point I would make about is big data are number 1, it is going to be absolutely critical that we can integrate these data together in means that are appropriate to create predicted models that can scope health and wellness for each individual.

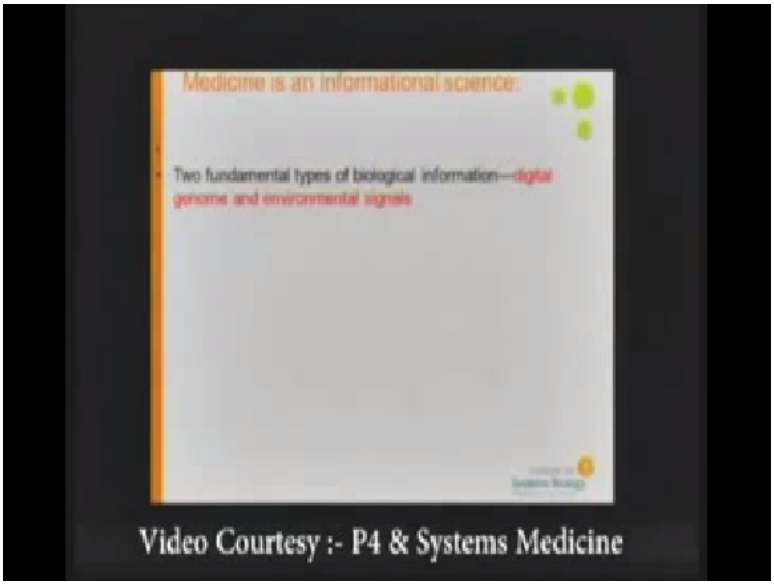
And number 2, we have to realise that the challenge for all big data is the signal to noise problem and I am going to emphasise the noise problem cannot be solved alone by machine learning, rather you have to integrate with machinery enormous domain expertise and I will show you an example of that.

(Refer Slide Time: 10:02)



So systems medicine deals with this complexity by thinking about biology as an informational science by creating an appropriate infrastructure for generating affectively larger amounts of data by creating the experimental approaches that can be holistic in nature and then having each of those drive the emerging technologies that open to various data space and their corresponding analytic tools. So let me make 3 points about medicine as an informational science.

(Refer Slide Time: 10:31)

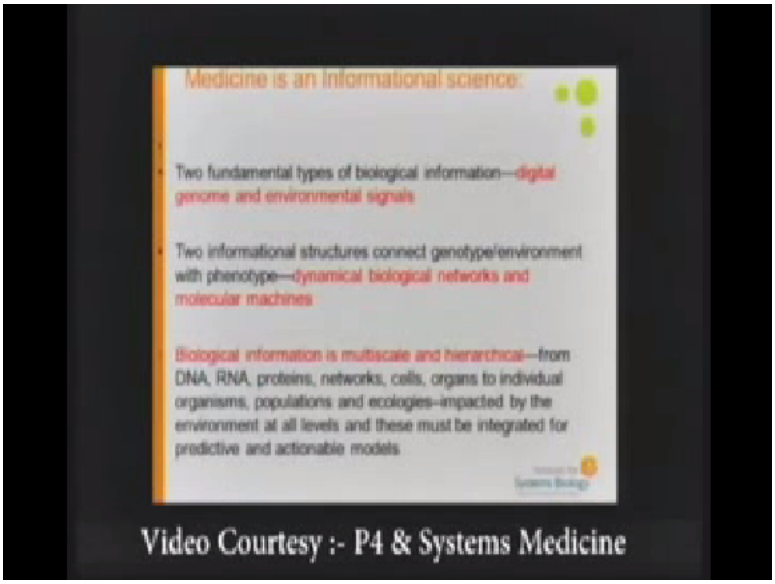


Number 1, there are fundamentally 2 types of data in living organisms, the digital genome and the environmental signals would come from the outside, together they blend this information to create phenotype of the organism and of course what connects the types of information with phenotype are 2 informational structures biological networks that capture and transmit, integrate

and finally pass off to the molecular machines that execute functions of life, the information that creates phenotype.

And what is the central focus of systems approaches are the dynamics of these 2 types of informational structure.

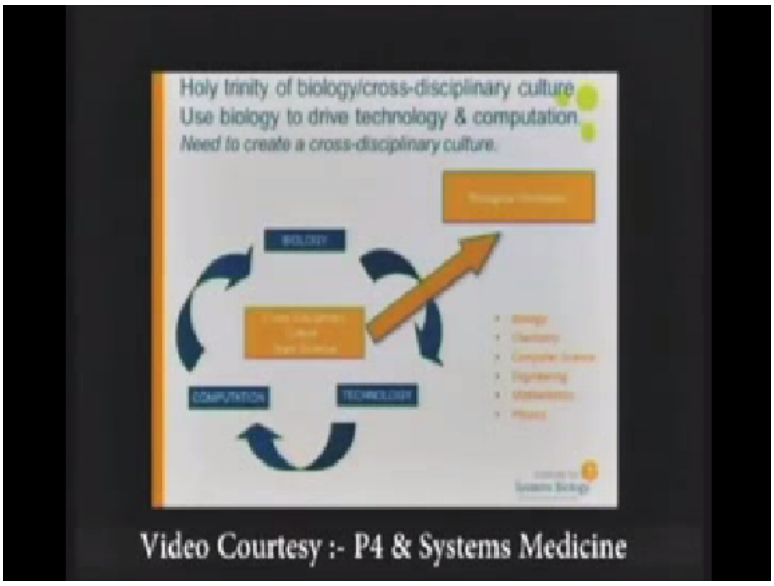
(Refer Slide Time: 11:19)



And the other point that I would make is that biological information is obviously multiscalar, DNA to RNA to proteins to networks to cells, all the way up to organisms, individuals, populations and acologies and the critical thing is that each of those levels, the environment imposes new signals on how the information is used and accordingly what is necessary is to capture as many of these levels of information as possible.

And integrate them in a manner to explicate some nature of the environmental signals for to completely understand the complexity of systems. You to understand the digital input and as well the environmental input. How you do these integrations is one of the grand challenges of systems biology. I would argue one of the most powerful aspects of systems approach are the appropriate infrastructures that are created.

(Refer Slide Time: 12:18)



When we started the institute for systems biology 12 years ago, the mantra rehab was needing edge biology mandates the need to develop new technologies, they in turn mandate the necessity for new analytic tools and that the biology seamlessly drives technology and computation in ways that can revolutionize the biology. To make that work, you have to have a completely cross disciplinary environment with.

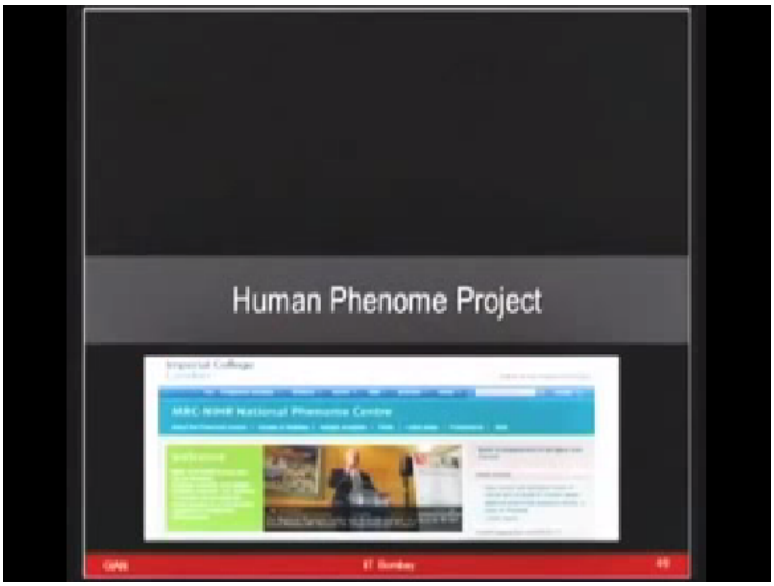
What are the most important or interesting facts from his talk and of course his work is that not only one need to measure the problems happening from the healthy versus diseased, but what is happening in the same healthy individual as well looking at your wellness profile, so the new project which they have initiated is on the same individual look at entire OMICS profiling, look into the genome to the transcriptome, proteome and metabolome.

Integrate those information of the same individual at different time points and then you can try to predict if there is everything happening okay in the body or there is some disease could be indicated, some pathways might be changing to indicative of a disease might be prone and that is a wellness study. So in which way P4 medicine can transform and the medicine actually is really being led by this group and their center and now there are certain companies which have come forward to take the P4 medicine for the wellness studies.

"Professor - student conversation starts" (()) (13:47) still it is personalized approach, yes but

how it would do the systems network and generate that information coming out on the OMICS data was the key from this work. **"Professor - student conversation ends".**

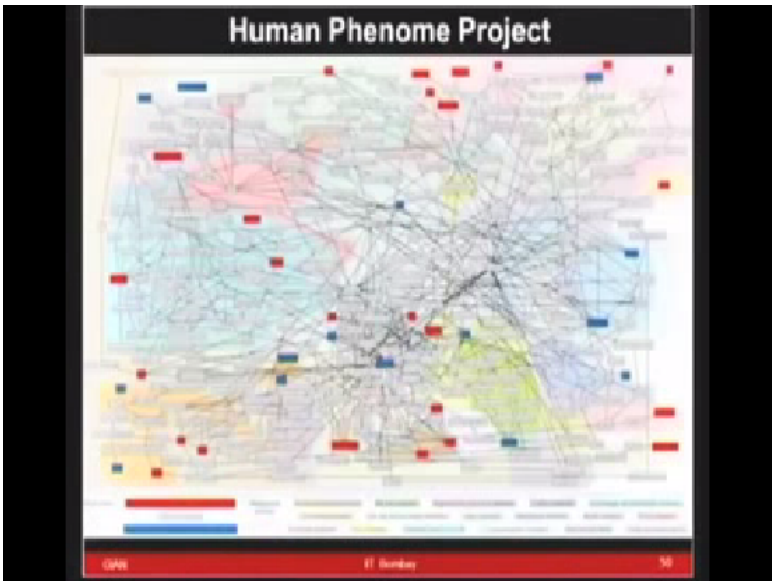
(Refer Slide Time: 14:02)



Another major initiative happening looking at the metabolites of individual and this project is happening at the Imperial College, London on the human phenome project. In fact, the leaders of this project were recently in IIT, Bombay for this workshop. Now you have talked to know, we talked about precision medicine on the genome level. I told you how to integrate those using the P4 medicine.

This initiative is happening at only the metabolomic level, large number of metabolites being measured on hundreds of thousands of individual using both mass spectrometry and NMR kind the platforms to look into their serum, their urines, saliva, all type of body fluids and their interaction with the environment.

(Refer Slide Time: 14:46)



So that is the human phenome project which was one of the mega funded project from UK government along with some support from London Olympics what they obtained and the 10 million pound grands. Now this is a real at high throughput level this project happening with several mass spec and several NMRs to measure the various metabolites changes happening in individuals.

And Dr. Jeremy Nicolson and his group from human phenome project recently published one paper in the science translational medicine where they have shown the entire network being modulated in individuals and they measured large number of patients for the urine samples. Again these kind of information is being generated now linked with the clinical information then with the environment factors is getting very important to provide the right cues.

Key object is the 21st century medicine is to understand an improved personalised healthcare and to understand and improve the population health as a whole and individuals and populations, the biology and also the likelihood of getting the disease is affected by the gene to environment interaction. On most gene environment interactions not only change individual probabilities getting disease, but also they change your metabolism.

So the metabolic phenotype effectively is where genes and environment meet. So it gives you a readout on some bodies biology that is not just captured by the genome alone. So what we are

trying to do here at Imperial and have been for a number of years is apply a whole range of advanced technologies, things like nuclear magnetic resonance spectroscopy, mass spectrometry for metabolically profiling complex samples like the urine and plasma and tissues, biopsies and that sort of things.

And in order to get profiles that relate not only to disease process but also to disease risk factors and we have established a very large center here to do exactly that. Recently because of the Olympic games, there has been a drug testing facility that has been created to screen the athletes for illegal substances or relation to their sporting pursuits. The analytical technology and instrumentation is very similar to that was required for phenotyping on a large scale.

So we have been successful in putting together a proposal at Kings College, London for taking over the drug testing facility at the end of the Olympic games and to make this equipments to instrumentation available to the UK university population for doing large scale metabolic phenotyping. So the technology that we have developed here in Imperial, we will adapt and develop to solve this much wider use by the community.

The need of stressing is that all over the country many (17:42 - 17:43: voice not clear) have been connecting (()) (17:45) of patients with particular conditions or large scale population individuals with a very large number of bank samples and the genetic determines are those being addressed by (()) (18:00) another large scale starting and given to often of great interest but also shown a great deal of disease remains unexplained at the moment by genetic basis complex trends and (()) (18:15) when this new facility comes in now as to look very much in the effects of genes in the environment acting together by looking metabolic profiling.

And (()) (18:28) hopefully at very low cost for individual samples starting and we think this will address a very large number of percentage across the UK and more widely across Europe and North America though we think we should take the opportunity to use this resource together with the bioinformatics behind it and the scientific expertise that Jeremy Nicholson's group.

There are number of different outcomes in scientific and medical terms that come from this type

of things having activity. The first one is to improve on knowledge about population health in general by metabolic and classifying individuals from large population studies and linking those into potential disease risk in the future and if we understand more about those gene interact by environment interactions that create health problems then we can legislate better and advice people better about their lifestyle management. So that is a major outcome of this sort of activity.

We also are going to be discovering new biomarkers that relate to disease processes and those biomarkers relate to mechanistic processes in cells and body which will give us insight into the disease itself and then potentially get new methods of attacking those disease, new therapeutic target. There is an another angle on this is to do with technology. So we have important strategic partners Waters Cooperation in mass spectrometry.

(()) (20:00) in nuclear magnetic resonance and these partners are putting in a lot of investment into developing the next generation that analytical technology and to have miniaturization, increase use of green solvents because when you do large scale analysis of these sort, you use a lot of chemicals and we want to make that more financially efficient and also better for the environment and also all the technologies that we are developing here.

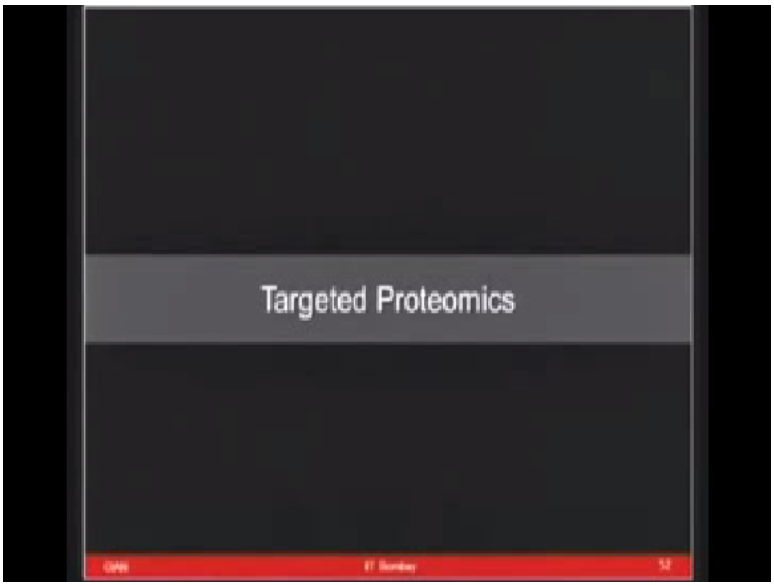
And the computational methods for extracting information are highly relevant to other areas of analytical science related to the environmental analytical research pesticides, pollutants, food science. The list is really quite comprehensive. So in addition to addressing a series of medical needs and problems solving to do with population health, there is going to be a long term legacy of this work in terms of analytical science for the United Kingdom.

Ideally this study has started showing its kind of bigger impact because initially it takes lot of investment to do this kind of large profiling, it takes very dedicated huge servers to save this data, just imagine that same sample being analyzed from NMR and from mass specs and then how much several TV data you are generating from each experiment, storing those having the right databases.

And linking those with the clinical information with those I think they have really started

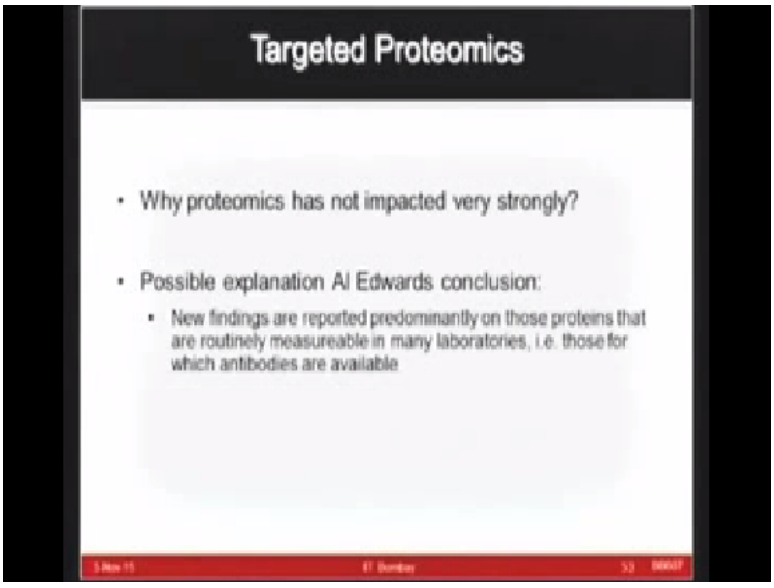
reaching out to the right information now and several series of very good papers that are coming out from this group and this work and their intention is also spread out these kind of human phenome centers in different parts of the world just so that this kind of technology could be utilised for doing metabolic profiling.

(Refer Slide Time: 21:44)



Target proteomics is an another revolution in field in progress happening. It was known as the nature methods of the year 2012 so the technology has just kind of emerging, 3 years old.

(Refer Slide Time: 21:58)



And in this area which you have seen that the major question of proteomics critics has been that why proteomics, although it is giving lot of data, but it is not impacting the clinical research

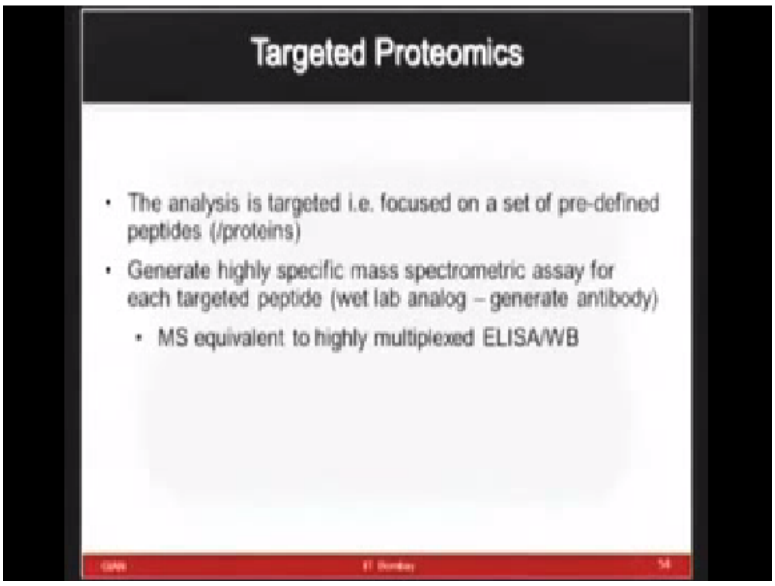
translational research. What is the reason for that? So from one of the papers, Dr. Al Edwards, what he makes conclusion which is kind of very good summary of why proteomics has not made huge impact on the field is.

We have always been trying to validate the proteins for which we have antibodies available and antibodies people have, companies have raised for those targets which are highly abundant, right and for very few regarded. So whenever you need to validate your proteins, you are always linked to the antibodies available commercially or in the market and so always you are going to look for haptoglobin, you are looking for hemopexin, you are going to look for those proteins for which you have got antibodies available.

Whereas in your mass spec data, you will have many proteins which are kind of very unknown, which are showing you huge (()) (22:54) changes but you have no way to do validation for those and therefore those proteins die because you are not able to validate and till you validate, nobody is going to rely on your data. Therefore, in absence of target proteomics field, this actually aims to look into the mass spec data further and do validation from the peptide level information.

So far the field has been relying on the antibody based assays, we talked about even histochemistry, we talked about ELISA, you talk about Western Blots, all of these are depending on antibodies, good antibodies. So the speed has not made major impact because there were no good validation strategies available and what is likely mentioned here new findings are reported predominantly on those proteins that are routinely measured in the minilab for which good antibodies are available. So that has been a major limitation why proteomics has not made impact as strongly (()) (23:44) it should have.

(Refer Slide Time: 23:44)



Now with this target proteomics approach, you can have the set of predefined peptides which are very abundant in a given disease for which you considered them as a potential biomarker. Now you can take those and generate very specific mass spec-based assays directly for those. So we were doing discovery and quantitation in the mass spec and you are also doing validation using mass spec. There is no role of antibodies in between and therefore what you discover, you are validating that.

Otherwise, you have discovered something else, you have to generate some new reasons for those, then you do validation independently and many times it fails into that pipeline. So many of the good hits and good targets which people had, now there is another opportunity for them to start validating and in fact, it has started making good impact and couple of markers are now being approved from US FDA and other regular key agencies. So therefore, there is a good hope that this field is going to make impact.

Improving human health by unlocking the mysteries of diseases is a goal that many scientists share and it will be enabled by mapping the proteins with a human proteome. Advancements in proteomics analysis are offering people new hope for the development of more effective drugs and the promise of protein biomarkers for a more personalized approach to medicine.

At the Swiss Federal Institute of Technology known as ETH Zurich in collaboration with the

Institute for Systems Biology, scientists have achieved one of the most significant advancements in the proteomics field to date generating a mass spectrometry assays for each of the proteins that correspond to the 20,300 genes in the human genome.

With the MRMAtlas project, we try to basically provide resource through the scientific community so that we develop these assays once and for all and then people can download these assays. The ultimate goal is to apply proteomics to biological and clinical research. We try to understand how biological processes function and are controlled by measuring the proteome.

Importantly the MRMAtlas expands in some never before seen areas of the proteomes. Scientist will now be able to analyze the dramatically more proteins in the human body with greater specificity, speed, and reproducibility in ever previously possible. The completion of the MRMAtlas marks the important turning point in the field that is expected to affect the biomedical research for years to come.

We have been now for a few years working very hard from developing and introducing targeted mass spectrometry. It is gaining (()) (26:36) a lot of attention of attraction getting very widely used. The technique that supports it is referred to selective reaction monitoring also frequently if it was multiple reaction monitoring. In contrast to immunoassays, we do not need to inject any animals.

But we need to generate (26:57 - 27:05: voice not clear) technology is an essential thought for us to support targeted mass (()) (27:06 - 27:11: voice not clear) instrument is the only instrument that can generate full fragment ion spectrum in a very (()) (27:17) high sensitivity. and as well as then use these (27:25 - 27:29: voice not clear) so in that sense, it is an ideal instrument for us because we spent lot of time developing large scale assays.

And then to use them for the measurement it can do this alone on the same platform. AB SCIEX has consistently produced high quality innovative instrumentation, also chemical reagents that are useful in the context of proteomics which is expanding and will be driven by new questions that can be asked and AB SCIEX will be there to deliver the answers.

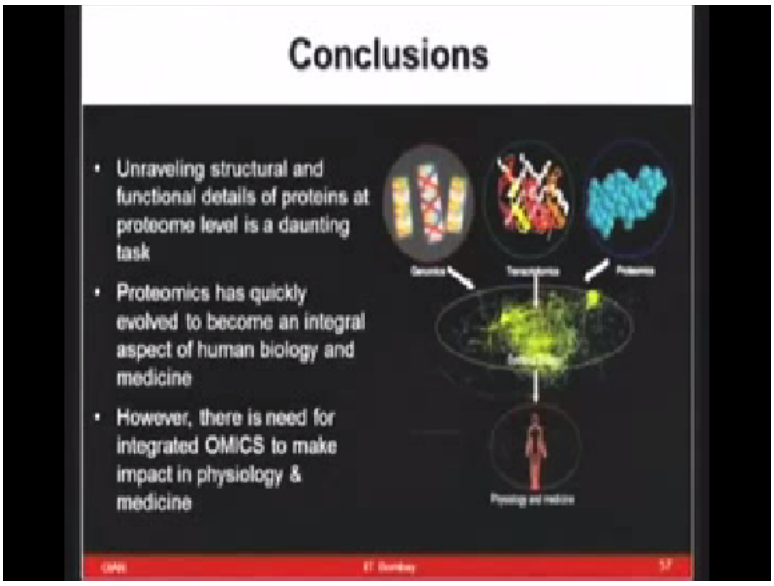
Alright, so Rudy Eberstadt is one of the pioneer in the field at ETH Zuric and his diverse kind of instrument that are developing (()) (28:20) MRM assays and so (()) (28:25), Dr. Robert (28:26 - 28:32: voice not clear) skyline and with all of these developments happening parallel now, this is looks like a new way this field is progressing now where biomarkers can be validated at the protein level.

(Refer Slide Time: 28:44)

And in United States, National Cancer Institute, NCI, they have now launched this new project from CPTAC which is Clinical Proteomic Tumor Analysis Consortium, CPTAC, they have now launched this precision proteomics project where these kind of (()) (28:57) being developed for various proteins. So these transition information for various peptides, you can obtain download from there server and then you can utilize on your own samples and you can do the validation.

So they have a dedicated team of the scientists and the software analyst who are already developing these assays now, the way you used to get the kits by ELISA, now you can get the information for the (()) (29:19) for the given proteins and you still need to adjust the (()) (29:23) and you need to adjust the conditions to do validation but now things are much more robust because they have a dedicated project in place where you can obtain lot of clinical assay information from the (()) (29:34).

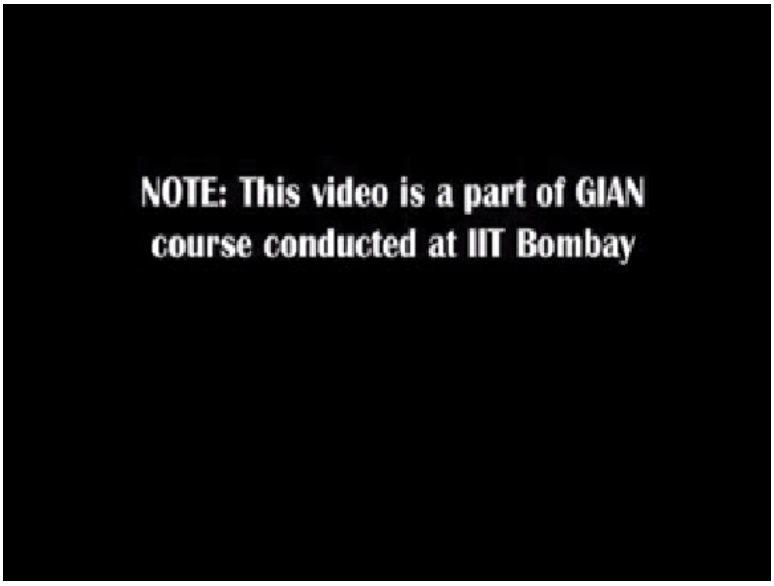
(Refer Slide Time: 29:36)



Well so in conclusion, in general lot of changes at the genome transcriptome proteomics is giving rise (()) (29:43) biology information which is going to be informative and linked for the physiology and medicine, while we have been looking at the proteome-level information which have been very doubting, very challenging, being very dynamic and very low-abundant proteins relevant for diseases.

Proteomics in general has evolved quickly to make an impact not only in the physiology and medicine but in variety of other application and all over the life sciences and biology in general. However, until unless we integrate this protein information with other OMICS information, I do not think we can really unravel of these complexity at the real biology level. So we need to be very open to integrate information from other OMICS level as well. So that is kind of conclusion from this presentation.

(Refer Slide Time: 30:28)


NOTE: This video is a part of GIAN
course conducted at IIT Bombay