

Proteomics: Principles and Techniques
Prof. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay

Lecture No. # 08
Proteomics and Systems Biology

(Refer Slide Time: 00:34)

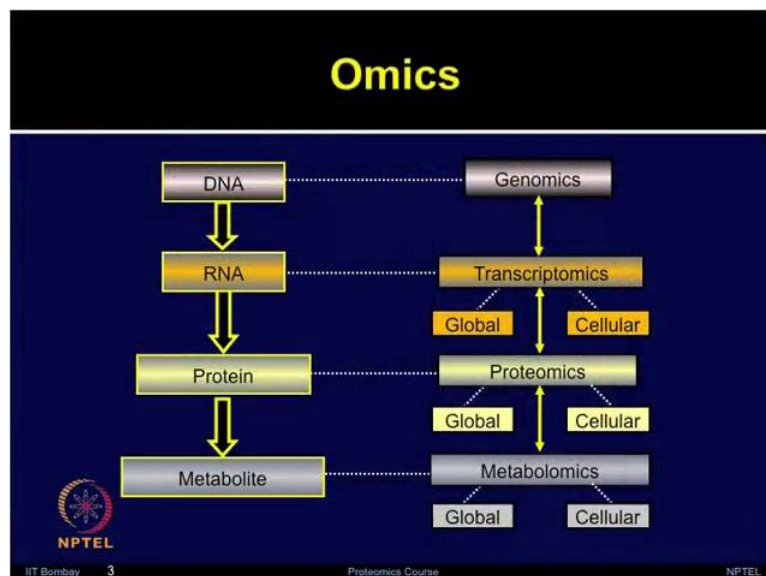
Lecture outline

- Proteomics
- Systems Biology

NPTEL
IIT Bombay 2 Proteomics Course NPTEL

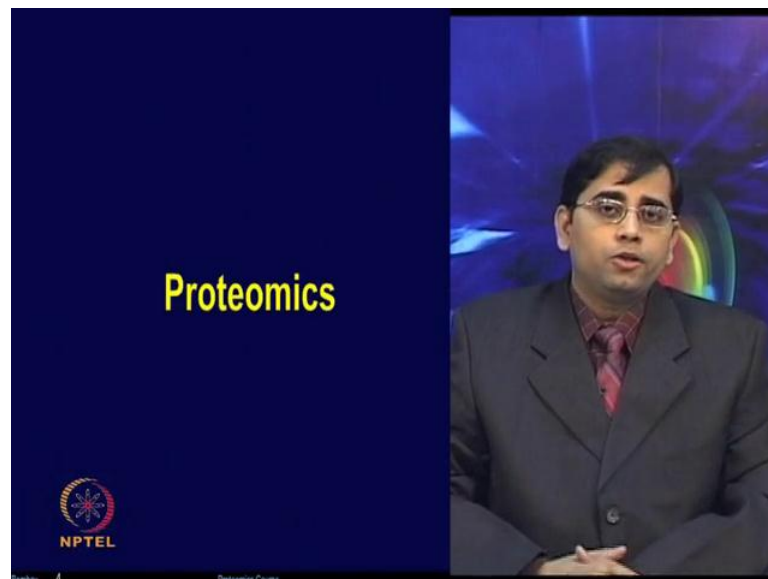
Welcome to the proteomics course, today's lecture is on proteomics and system biology, the lecture outline; first I will start on proteomics, and then we will move on to the system biology.

(Refer Slide Time: 00:55)



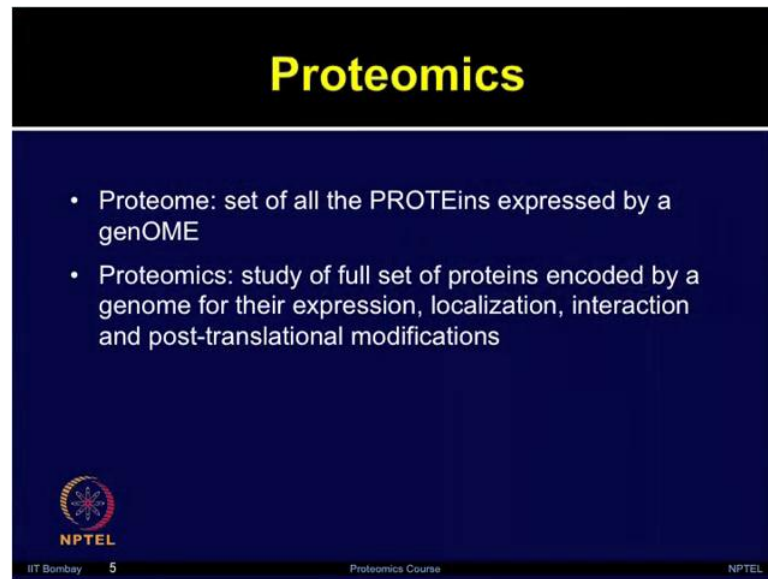
Today we will continue on first proteomics, and then how different type of omic data can be applied for systems approach analysis that will be discussed. So, this slide I have shown you how you can define different type of omic platform, so omic is a suffix derived from the Greek word om means all or every, the omics is used is a suffix which has enabled the explosion of terms genomics, transcriptomics, proteomics and metabolomics and so on and, so for omics also implies integration of biology with information science and conveys large scale biology using systems approach, so as you can see the slide if you are studying about DNA that will be in totality known as genomic, RNA study and transcriptomics, proteins and proteomics, and metabolites in metabolomics, if you are looking at all of the cellular contents of proteome that will be known as cellular proteomics or cellular genomics; and similarly, all of the proteome of an organism will be known as global proteomics or similarly at the gene level global genomics.

(Refer Slide Time: 02:19)




So, let us first start with proteomics.

(Refer Slide Time: 02:24)



Proteomics

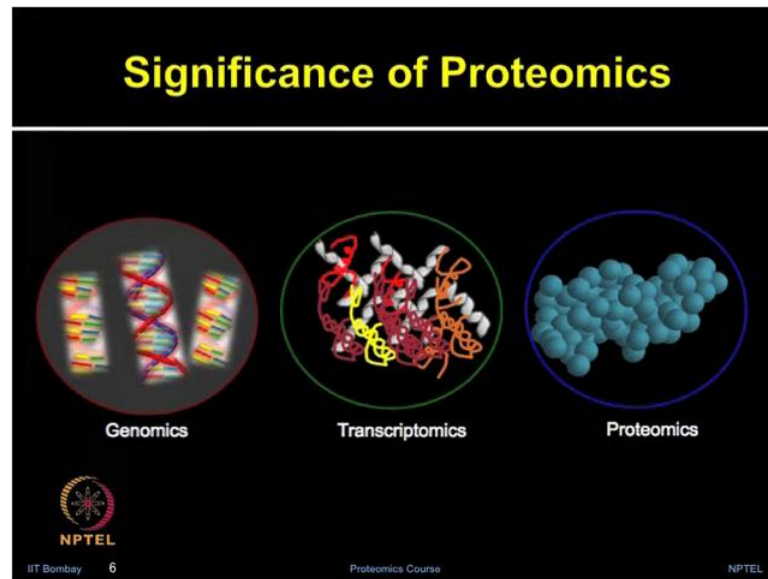
- Proteome: set of all the PROTEins expressed by a genOME
- Proteomics: study of full set of proteins encoded by a genome for their expression, localization, interaction and post-translational modifications


NPTEL

IT Bombay 5 Proteomics Course NPTEL

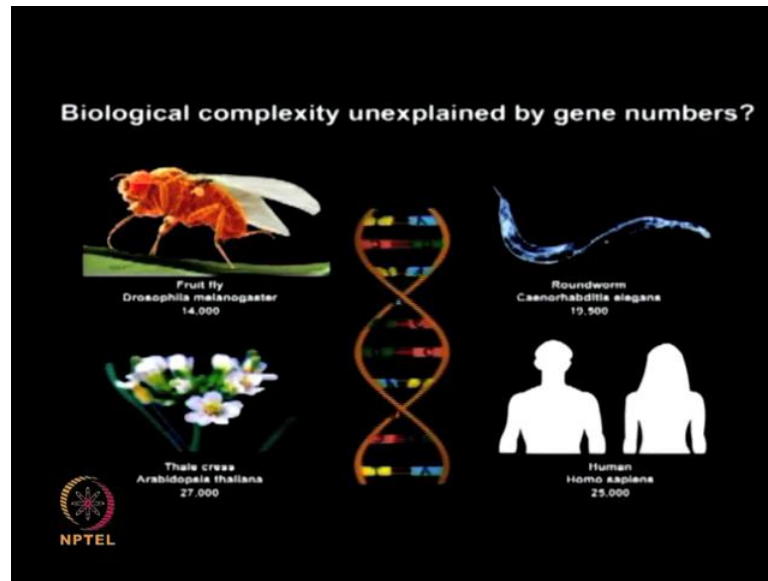
Say, if you remember from the previous lecture proteome is set of all the proteins which are expressed by a genome, proteomics a study of proteins and their properties to provide an integrated view of cellular processes, what are these different properties? These properties include the extent of the protein expression how different type of post-translational modification and co-modification occurs in the cell, different type of enzyme regulation whether it is activation or inactivation and then different type of intermolecular protein protein interactions. The current goals of proteomics are very broad it is including the diverse properties of proteins which we have discussed in the earlier lectures looking at the (()) chemistry of amino acids and different levels of protein structures, so whether it is sequence, quantity, the state of modification, activity, interaction of proteins with other proteins and other biomolecules and subcellular distribution and a structure analysis all of these are broad goals of proteomics.

(Refer Slide Time: 03:56)



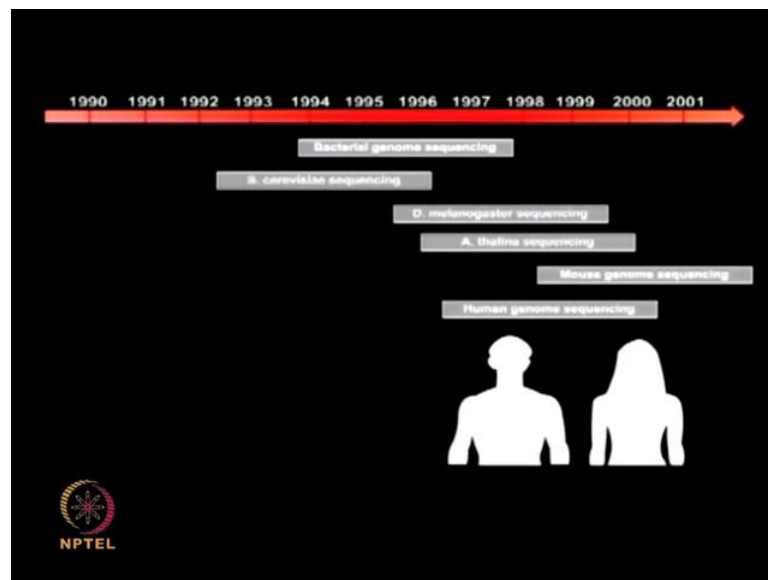
A central dogma concept was discussed earlier which is an orderly and unidirectional flow of information in coded in base sequence of cells passed on from DNA to RNA and then to the proteins this is the simplest definition of central dogma, the genome sequencing projects they have provided researchers with an unprecedented information of genome sequences; however, there are numerous proteins which can be in coded by the genome therefore, analysis of the static genome by doing sequencing alone is not sufficient, a gene can code for several type of proteins, because of it alternative splicing and post-transcriptional and post-translational modifications; therefore, it suggest that studying proteins is more challenging than genome or transcriptome; therefore, proteomics has great significance to understand the biological systems.

(Refer Slide Time: 05:13)



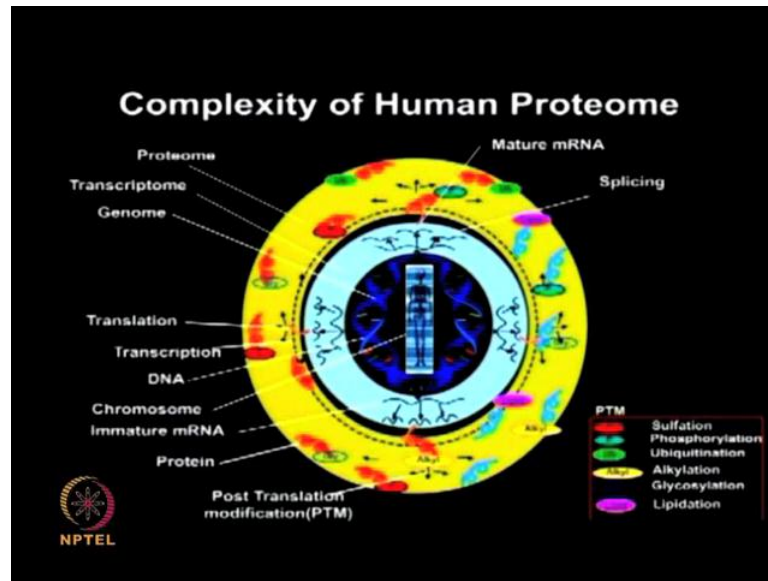
The complication of genome sequencing projects of several organism including human has been one of the most remarkable achievements of the century; however, these have not been sufficient to unravel the mystery of complex biological processes.

(Refer Slide Time: 05:34)

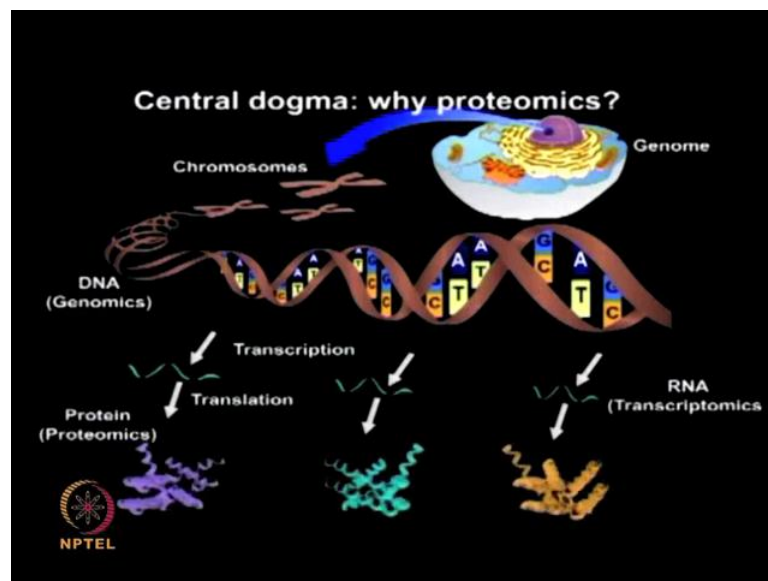


The similar gene numbers of mainly diverse group of organisms has failed to explain their varying biological complexity, a more meaningful understanding of biological function can be obtained through the characterization of products of gene expression, the protein which serve as ultimate effector molecule of biological systems.

(Refer Slide Time: 06:07)

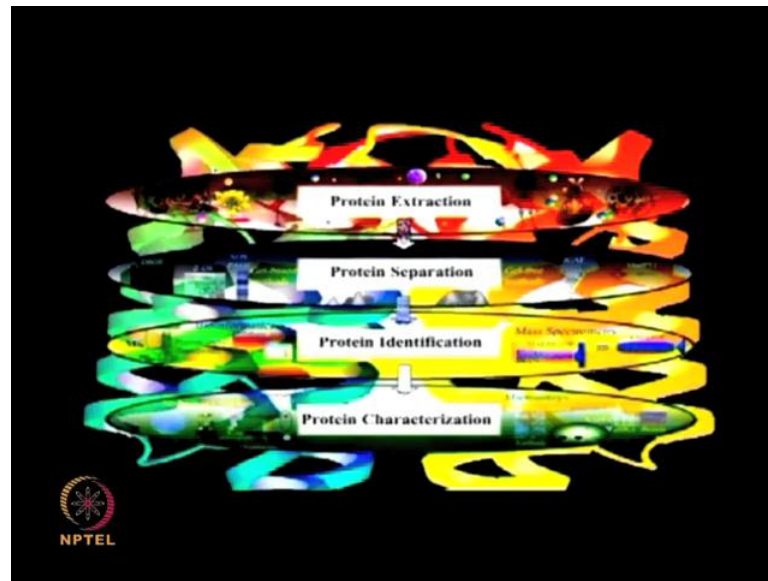


(Refer Slide Time: 06:51)



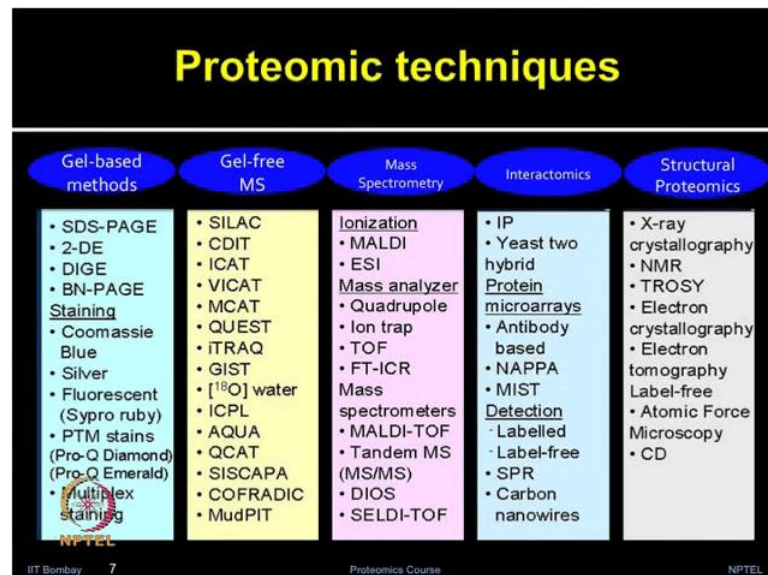
The proteomics refers to the study of entire protein complement expressed by an organism at any given time, while the genome of an organism is mostly static; the proteome is dynamic and it changes with environment and time; thereby, elevating it is complexity level, the gene regulation is regulated by several post-transcriptional and post-translational modifications due to which the number of proteins expressed in a cell is much greater than it is genomic counterpart.

(Refer Slide Time: 07:19)



The proteomics aims to decipher the structure and function of all proteins in a given cell under a specific conditions and to obtain a global view of cellular processes at the protein level, study either DNA level known as genomics, RNA level transcriptomics, and protein level proteomics.

(Refer Slide Time: 07:39)



Analysis of the proteome involves protein extraction, separation, identification and finally, characterization of various proteins, the various proteomic techniques which are currently employed in different applications, we will talk in detail about different

proteomic technologies in the subsequent modules of this course, but briefly very broadly you can group these technologies as gel-based methods, gel-free mass spectrometry-based methods, mass spectrometry techniques, techniques to study the protein interactions and structural proteomics, as I have shown some abbreviation in this slide broadly different type of gel-based proteomics such as sodium dodecyl sulfate, polyacrylamide gel electrophoresis, SDS page, two dimensional electrophoresis 2 d e, difference in gel electrophoresis DIGE, blue native page as well as different type of staining methods such as coomassie, silver, fluorescent dyes, and phosphorus stain such as process-q diamond and multiplex staining methods all these can be grouped under gel-based methods.

The gel-free methods especially in the mass spectrometry site includes SILAC which is a stable isotope labeling by amino acids in cell culture, CDIT culture derived isotope tags, ICAT isotope-coded affinity tagging, VICAT visible isotope-coded affinity tagging, MCAT is mass coded affinity tagging, and then QUEST which is quantitation using enhance signal tags, iTRAC isobaric tagging for relative and absolute quantitation, gist global internal standard technology, ICPL isotope-coded protein labeling, aqua absolute quantitation, SISCAPA stable isotope standards and capture by anti-peptide antibodies, COFRADIC combined fractional diagonal chromatography, and MudPIT which is multidimensional protein identification technology all of these are various new advancement in the gel-free methodologies.

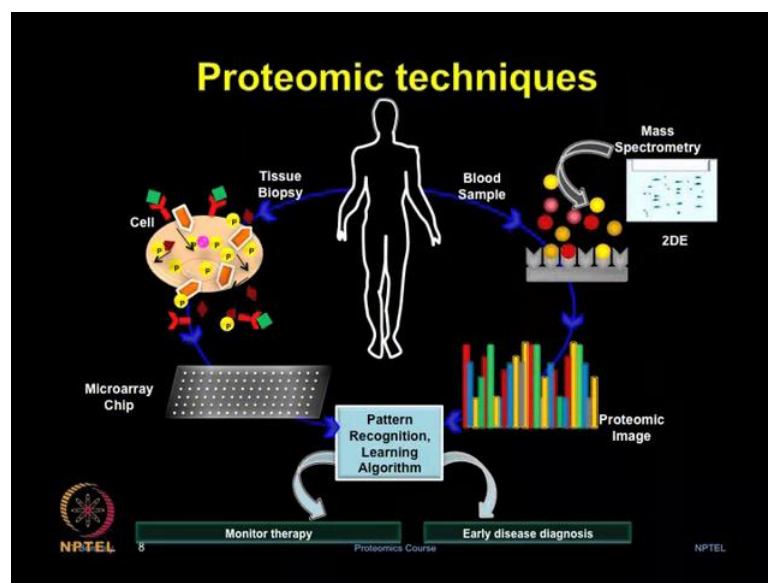
The basic mass spectrometry which is central to the proteomics application includes different type of ionization sources such as matrix-assisted laser desorption ionization maldi, electrospray ionization ESI, and different type of mass analyzers such as quadrupole, time of flight, ion trap and fourier transform, mass spectrometry then different type of tandem m s based systems are also used, the surface-enhanced laser desorption ionization time of flight seldi-tof is also used for various chemical applications.

The protein interaction methodologies include immunoprecipitation, yeast two hydride methods, and different type of protein microarray platforms such as antibody arrays, nucleic acid programmable protein arrays, multiple spotting techniques and various other cell based and cell free expression based protein microarrays, the detection can be either based on the label methods using fluorescence, chemiluminescence or radioactivity, or it

could be different type of label free methods such as surface plasmon resonance, interferometry based methods, or different type of conducted based methods employing nanotubes and nanowires.

The structural proteomics it involves X-ray crystallography, nuclear magnetic resonance N M R, transverse relaxation optimization spectroscopy TROSY, circular dichroism C D, and different type of microscopy methods including atomic force microscopy and electron microscopy.

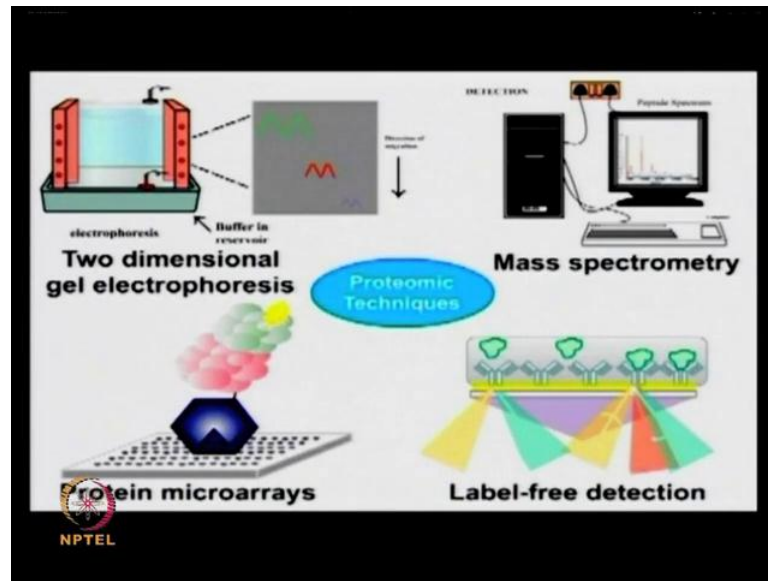
(Refer Slide Time: 12:47)



So, as we have seen the large number of proteomic technologies which are currently available for various applications, many times to address one biological question different type of methodologies come together and then provide solutions to that problem; for example, looking at some clinical sample for identification of biomarkers of a disease one can employ the samples such as tissue or blood or different type of body fluids and then either directly extract the protein and subject to the mass spectrometry, or first resolve on two dimensional electrophoresis followed by identification on mass spectrometry, or take these samples directly apply on the microarray-based platforms and then detect using label based or label-free methodologies.

Eventually these type of results will enhance knowledge for the monitoring the therapy response as well as identification of early disease diagnosis, this is just one example, similarly, multiple type of proteomic technologies can be used for different applications.

(Refer Slide Time: 13:49)




There are several proteomics techniques which are employed for studying these proteins such as two dimensional gel electrophoresis, mass spectrometry, protein microarrays as well as some label-free detection techniques such as surface plasmon resonance S P R.

(Refer Slide Time: 14:39)

Systems Biology

- Examination of a biological entity as an integrated system, rather than study of its individual characteristic reactions and components


NPTEL

IIT Bombay 10 Proteomics Course NPTEL

After discussing proteomics now let us talk about system biology, so what is system biology? Is that an esoteric knowledge, a method to understand biological systems or a tool to solve the practical problems, systems biology is the examination of a biological entity as an integrated system rather than the study of it is individual characteristic

reactions and components which is termed as systems biology, in study of all the mechanisms underline the complex biological processes in the form of integrated system of many interacting components is studied under system biology.

(Refer Slide Time: 15:11)

Systems Biology

- System level understanding of biological networks
- Common elements of systems biology
 - Networks
 - Modeling
 - Computation
 - Dynamic properties

DNA RNA Protein Systems Human

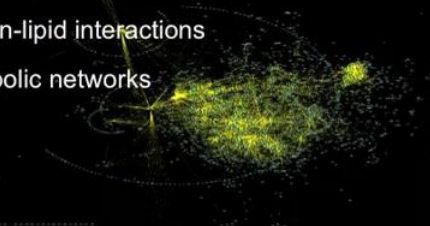
IIT Bombay 11 Proteomics Course NPTEL

The system level understanding of biological networks require information from different levels, as you can see from DNA to RNA to proteins for make systems and then that information can be applied to understand a complex system for different organism, the biological information is represented by the networks of interacting elements and dynamic responses to the perturbation.

(Refer Slide Time: 16:12)

Biological networks

- Protein-protein interaction networks
- Gene regulatory networks
- Protein-DNA interaction networks
- Protein-lipid interactions
- Metabolic networks

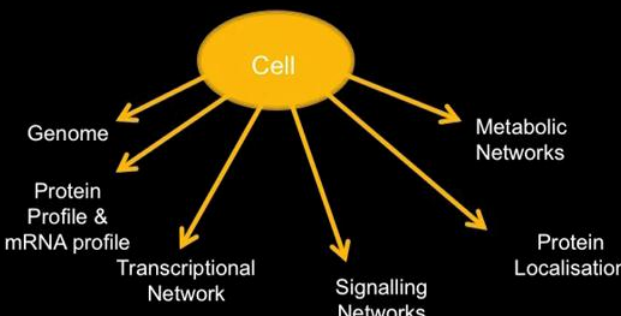


NPTEL
IIT Bombay 13 Proteomics Course NPTEL

These networks provide insights which cannot be analyzed from the isolated components of the system, the common elements of the system biology include networks, modeling, computation and dynamic properties. The different types of biological networks such as protein-protein interaction networks, gene regulator networks, protein and DNA interaction networks, protein lipid, protein other biomolecule networks and metabolic networks.

(Refer Slide Time: 16:26)

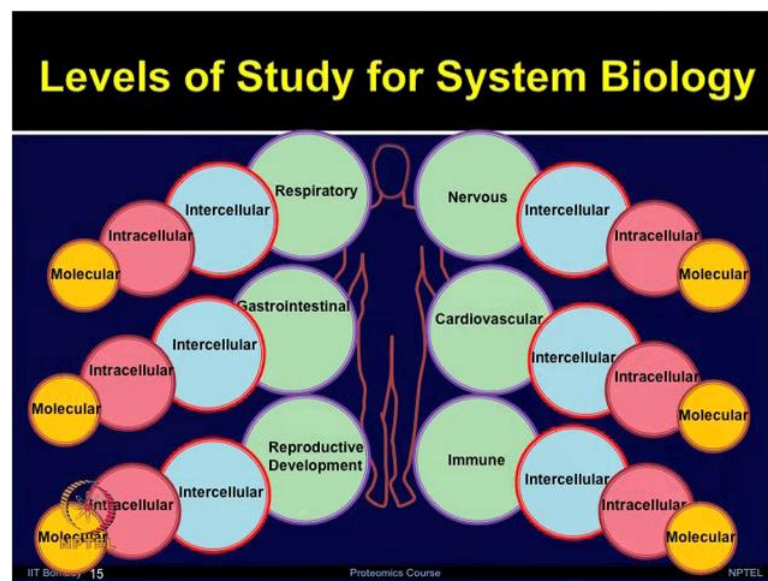
Ingredients of Systems Biology



NPTEL
IIT Bombay 14 Proteomics Course NPTEL

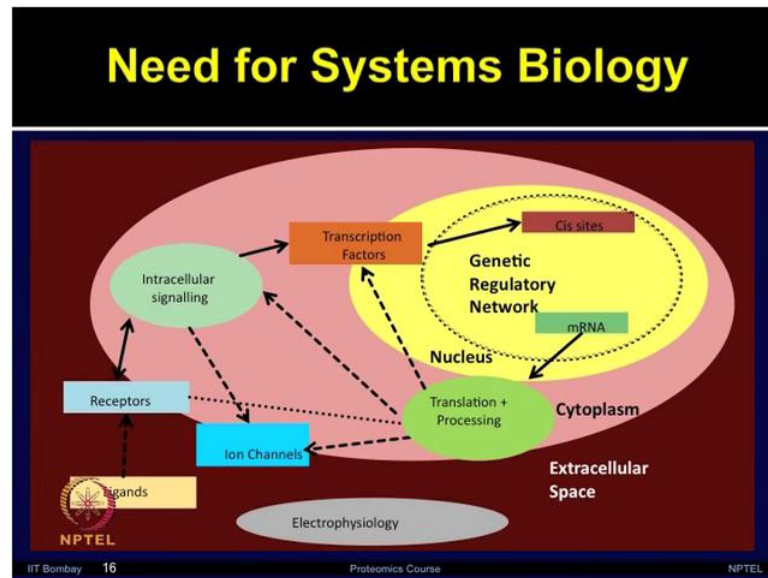
The various ingredients of systems biology for example, if you are studying about a cell and a system behavior one need to look at the genome it is transcriptome profile, proteome profile, how protein D N A and different type of transcriptional networks are altered, protein-protein signaling networks, multimeric complexes how they are found? How proteome is localized in the intracellular dynamist, and metabolic networks? So, all of these are ingredients of studying about a system.

(Refer Slide Time: 17:05)



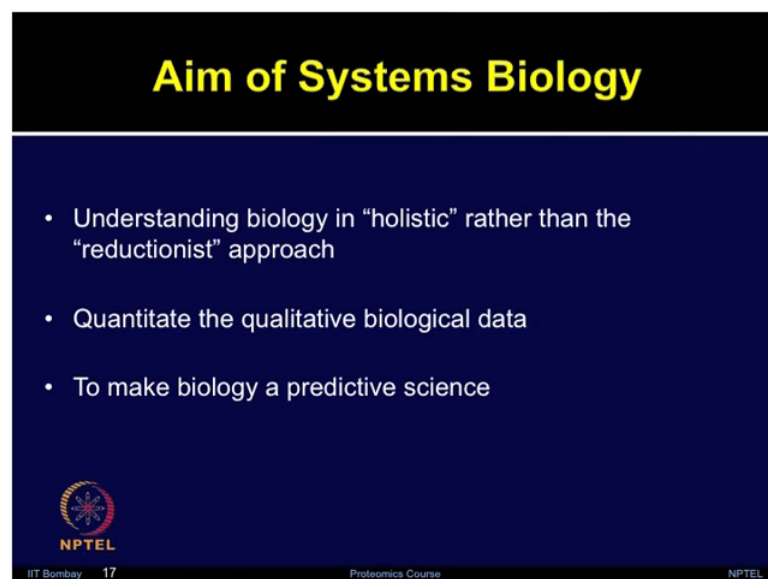
A systems biology study can be done at different levels for example, to study the complex physiology of human one can look at individual system such as respiratory system, nervous system or other physiological systems, studies can be done at the intercellular or intracellular level and finally, at the molecular level involving genomics transcriptomics and proteomics

(Refer Slide Time: 17:38)



So, why there is need for systems biology? The study of biology at the system and subsystem level for understanding the biological processes and network is very much required, as you can see to understand even simplest system of a cell how it is regulated with the it is extracellular space, the sitoplas and different other components, the examination of the structural and dynamics of cellular and organizable function is very much required for understanding the systems, rather than the characteristics of isolated parts of the cell or organism.

(Refer Slide Time: 18:40)




So, what is the aim of systems biology to understand the biology in holistic approach rather than the reductionist approach? The systems biology aims to quantate the qualitative biological data and provide some level of prediction by applying different type of computational modeling.

(Refer Slide Time: 18:53)

Systems Biology Approach

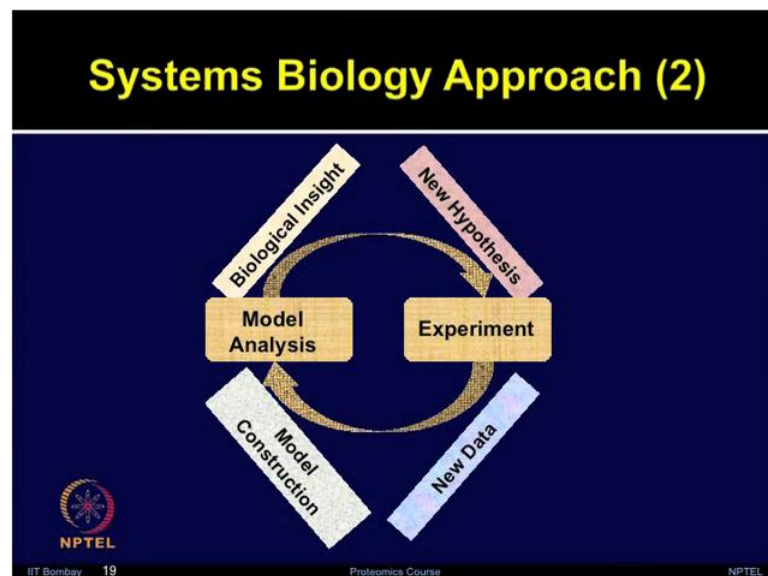
- Systems biology involves:
- Experimental data set collection
- Mathematical models



IIT Bombay 18 Proteomics Course NPTEL

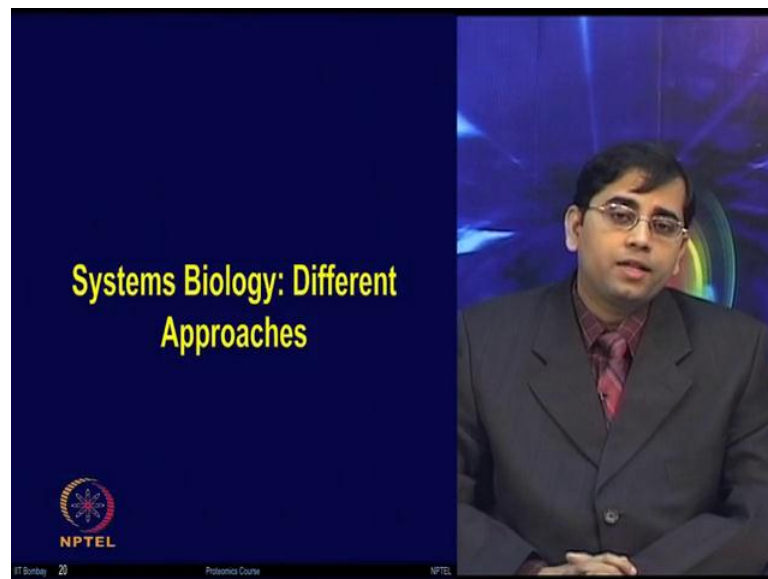
The systems biology approach involves first of all collection of large experimental data sets and then mathematical models to provide insight of some significant aspect of the data sets.

(Refer Slide Time: 19:17)



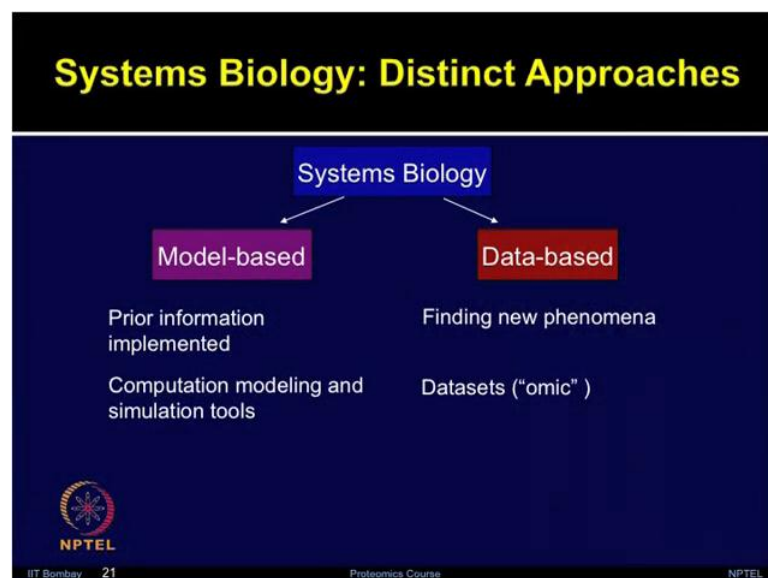
Simple systems biology approach would involve experiments by adding new data sets which will be used for model constructions and model analysis, and the biological insight derived from these models can be used to propose new hypothesis, so the properties of a systems are probably more than just the sum of all its individual properties or its components; therefore, it is possible that a system may have its own property by applying all the components.

(Refer Slide Time: 19:51)



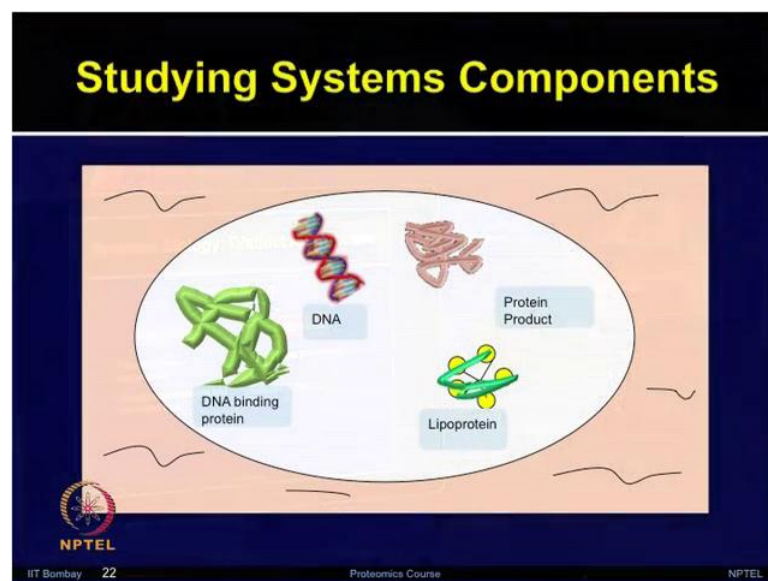
So, what are different approaches have been taken to study the systems biology?

(Refer Slide Time: 19:56)



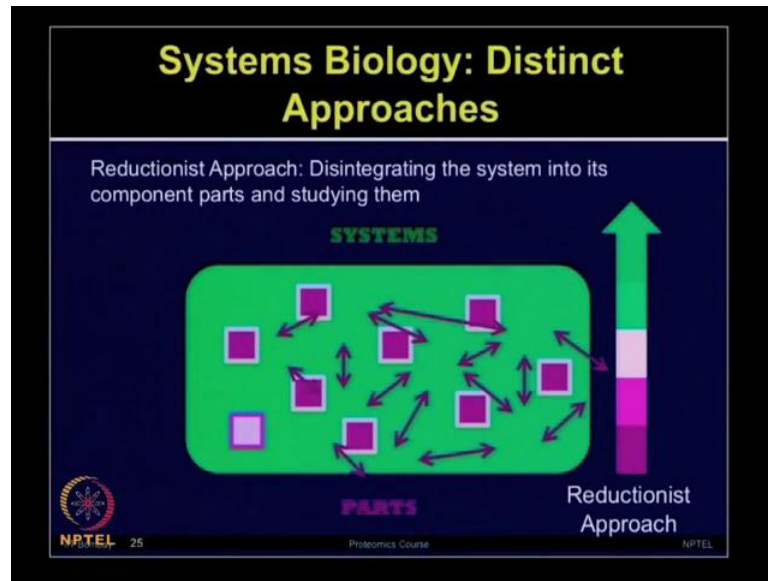
The distinct approaches of systems biology include the model-based and data-based methods, the model-based approach involves some prior information which can be implemented in the models; whereas, the data-based objective is to find the new phenomenon, the model-based relies on computational modeling and simulation tools; whereas, the data-based method relies on the omics data sets. So, in model-based systems biology approach it is difficult to build the detailed kinetic models, but in data-based system the complex relationship among the various type of omics information and metabolic pathways and networks can be created.

(Refer Slide Time: 20:51)



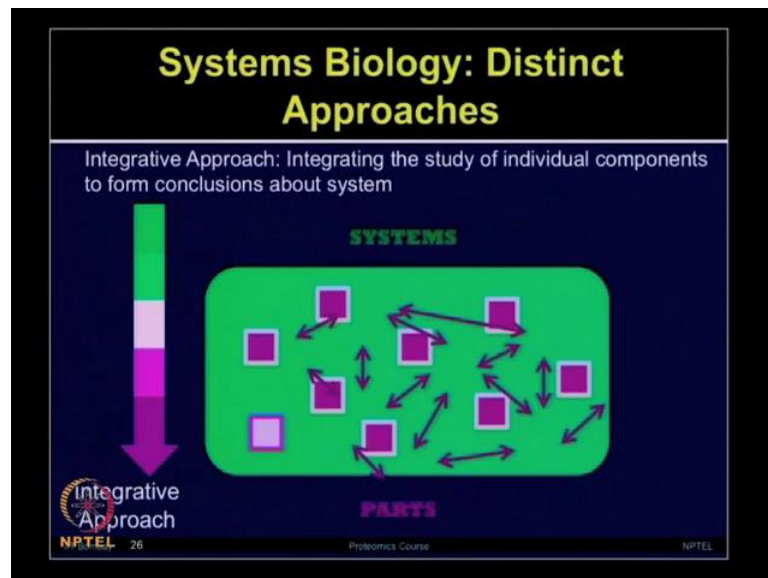
Studying systems component is very challenging, systems biology and biological network modeling aims to understand the systems structure and function for better understanding of system properties like it is robustness as well as use for the prediction of systems behavior in response to the perturbations.

(Refer Slide Time: 21:19)



The reductionist approach involves disintegrating the system into its component parts and studying them.

(Refer Slide Time: 21:29)



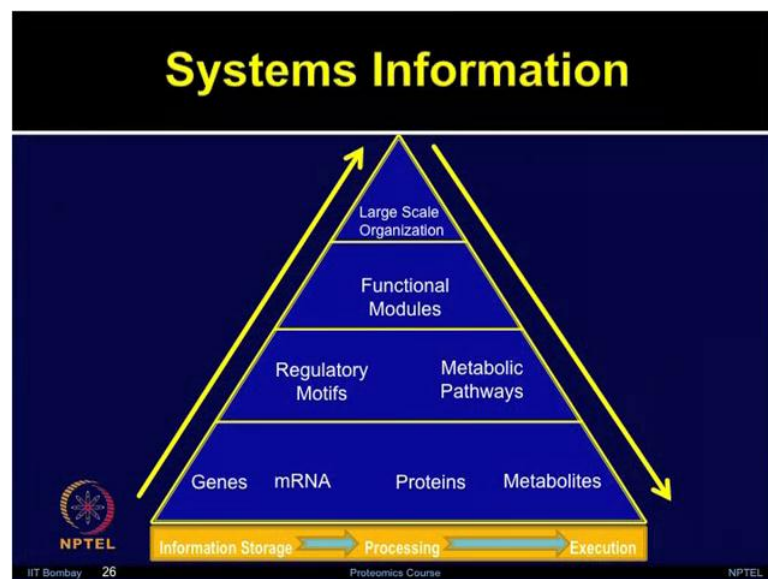
Whereas the integrative approach involves integrating the study of individual components to form conclusions about the system.

(Refer Slide Time: 21:41)



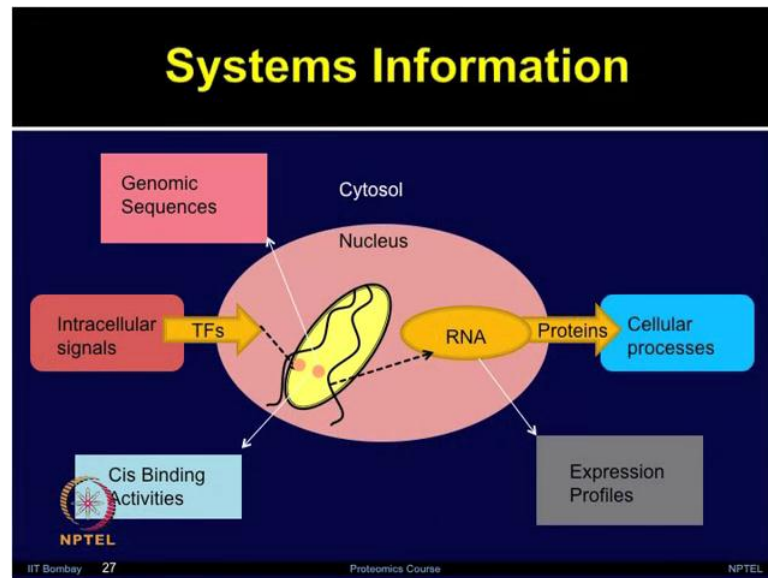
What is systems biology triangle?

(Refer Slide Time: 22:01)



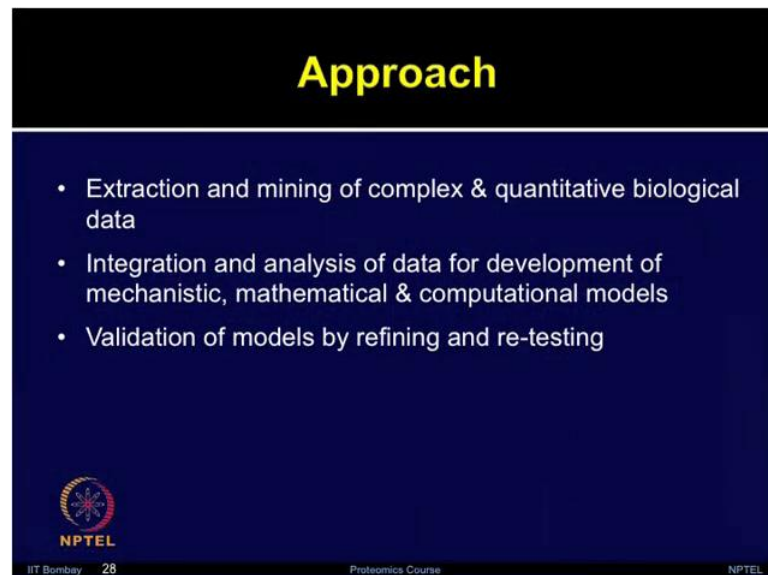
So, first of all the systems information is generated at various levels, as we have discussed starting from genes to mRNA to proteins to metabolite, or identifying regulatory motifs, metabolic pathways, functional modules and different large scale organizations, this information has to be stored processed and further executed to identify the system level information.

(Refer Slide Time: 22:28)



Even simplest system such as cell can be linked with various properties it is genome sequences of different molecules, intracellular signals transcription factors, different type of Cis binding activities, the expression profiling of RNA and proteins and different type of cellular processes.

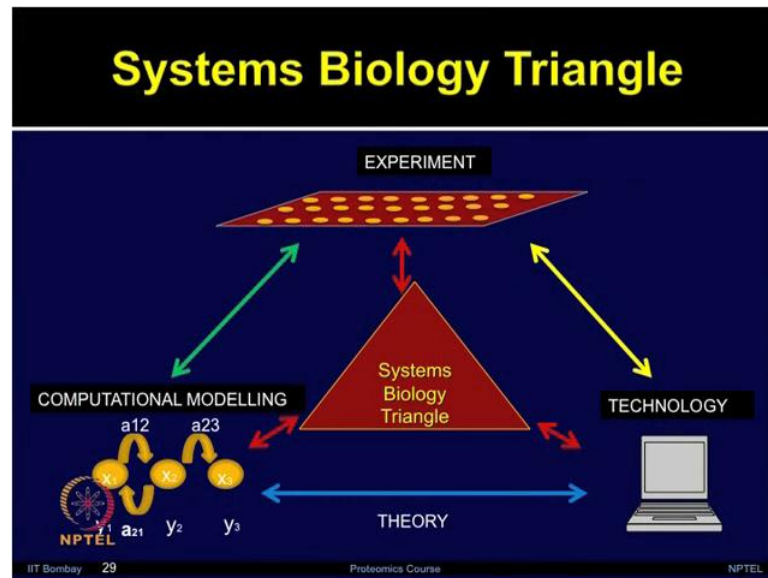
(Refer Slide Time: 22:51)



So, what approach one can take to study about the system? Extraction and mining of complex and quantitative biological data, integration and analysis of these data sets for the development of mechanistic, mathematical and computational models, validation of

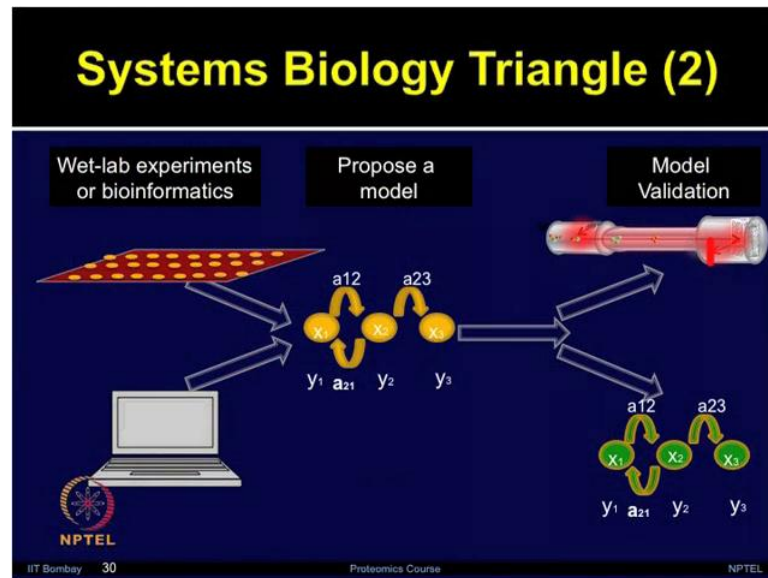
these models by retesting and refining after proposing from hypothesis, different online data bases and repositories are nowadays develop for sharing large data sets and various systems models, the systematic approach to study how molecules act together within the network of interaction that makeup life is definitely going to be useful to understand the systems biology.

(Refer Slide Time: 23:47)



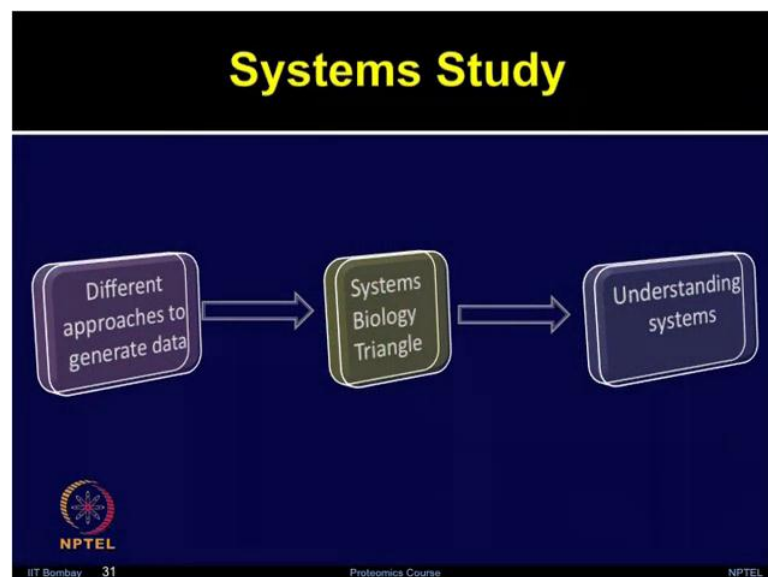
The systems biology triangle as you can see here involves the experimental data set could be derived from different type of omics platforms, technologies how the computational analysis can be performed, different type of bioinformatics software's and tools, and then computational modeling by obtaining sub theoretical concepts, this synergistic application of the experiment theory and technology with modeling to enhance the understanding of biological processes as whole system rather than the isolated part is termed as systems biology triangle.

(Refer Slide Time: 24:30)



So, the systems biology triangle the wet lab experiments or bioinformatics based data analysis can be used to propose a model, the model building as an aid to understand the complex system, and some hypothesis can be generated which could be used further to propose more quantitative models or predictive models, and also it can be used by independent techniques for the model validation.

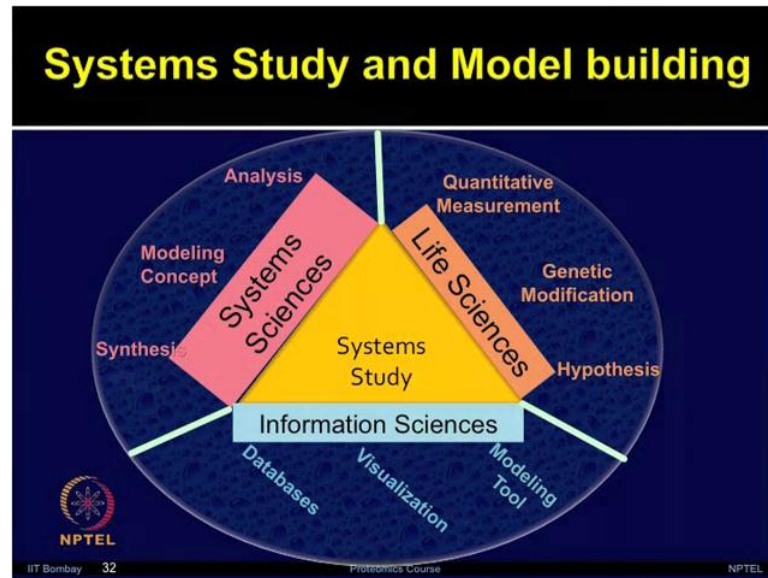
(Refer Slide Time: 25:02)



So, what is systems study? First of all the difference between systems study and the component study one need to understand, and this what we have try to emphasize in the

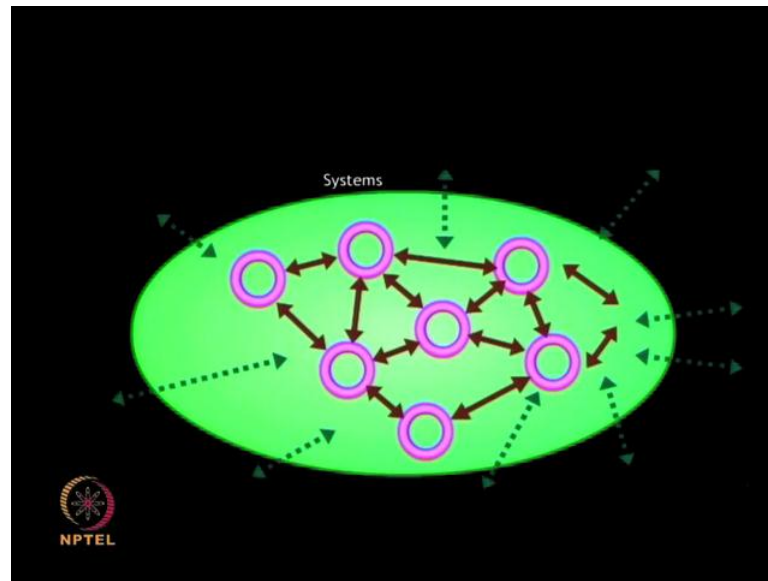
previous slides after generating the data set and creating the systems biology triangle then this information can be used for understanding the systems in the more complex and mechanistic level.

(Refer Slide Time: 25:32)



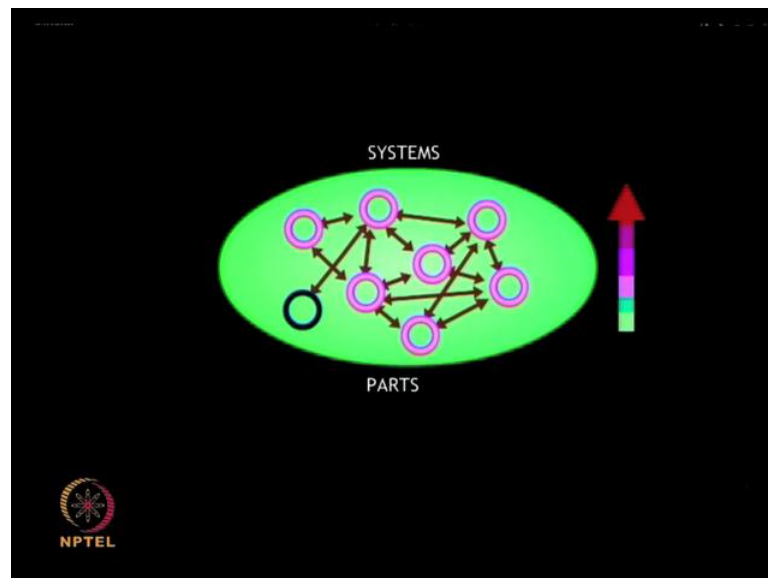
Systems study and model building, the system science includes synthesis modeling concepts analysis, life sciences discipline provides quantitative measurements, genetic modifications and deriving some hypothesis, then formation sciences enables the visualization, the modeling tools and different data bases, so this model building as an aid to understand the complex system is very useful for systems level investigation.

(Refer Slide Time: 26:10)



System is an entity which maintains its existence through a mutual interaction of its constituent's parts.

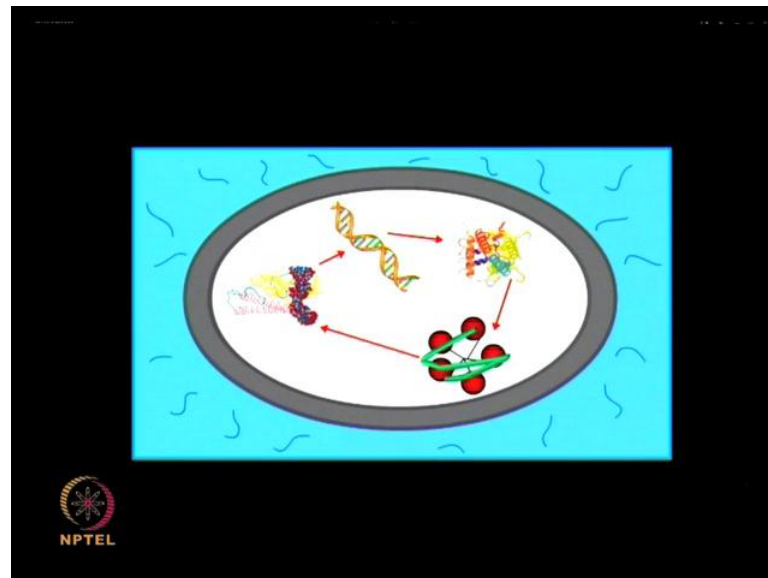
(Refer Slide Time: 24:56)



The systems biology research consists of identification of the parts, characterization of the components, exclusion of those which are not a part of the system, identification of the interaction of the components with each other, and identification of the interaction of the components with the environment which modulate the parts either directly or indirectly through modulation of internal interactions.

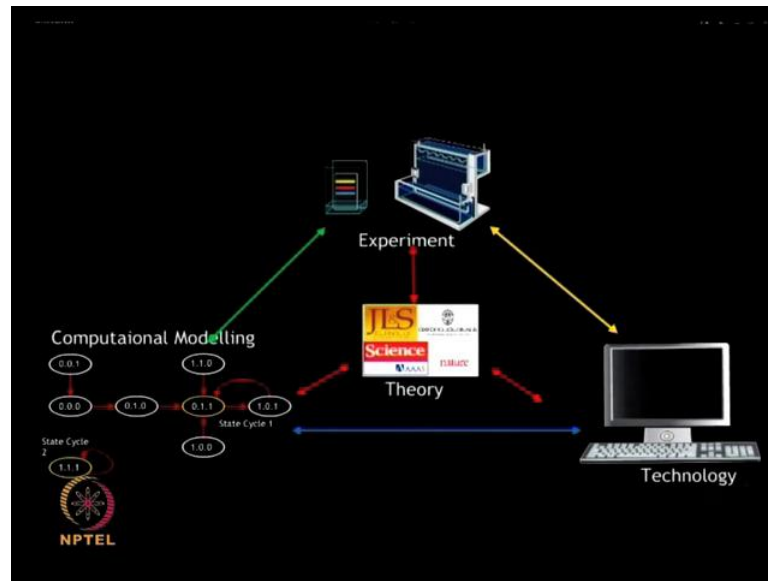
The system biology concept can be understood with help of two approaches such as reductionist approach and integrative approach, the reductionist approach focuses on disintegrating the system into its component parts and studying them; whereas, the integrative approach focuses on integrating the study of individual component to form conclusions about the system.

(Refer Slide Time: 28:23)



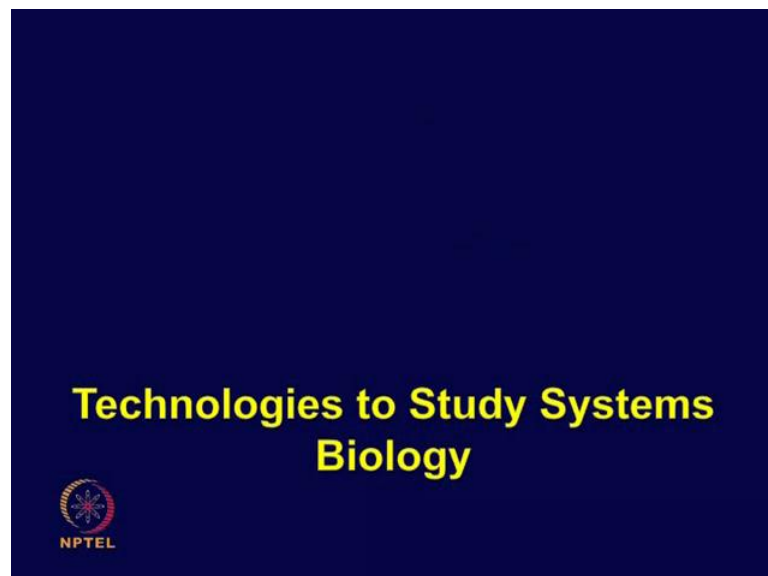
Consider a cell with its component molecules, let us (()) study the metabolic pathway as a biological system, when the environment of this cell is perturbed a little the individual components undergo unique changes such as increase in the production rate or decrease in their amount, at this stage due to lack of knowledge of the nature of interaction of proteins we cannot interpret how the system gets affected, but when we study the interaction of one component with the other we can conclude that the increase in rate if DNA binding protein leads to increase in the synthesized amount of DNA which further changes, the final amount of lipoproteins produced thus we can see that we study a system we need to analyze not just the components but their interactions, these biological systems can be protein-protein interaction networks, gene regulatory networks, protein DNA networks, protein lipid networks and metabolic networks.

(Refer Slide Time: 29:31)



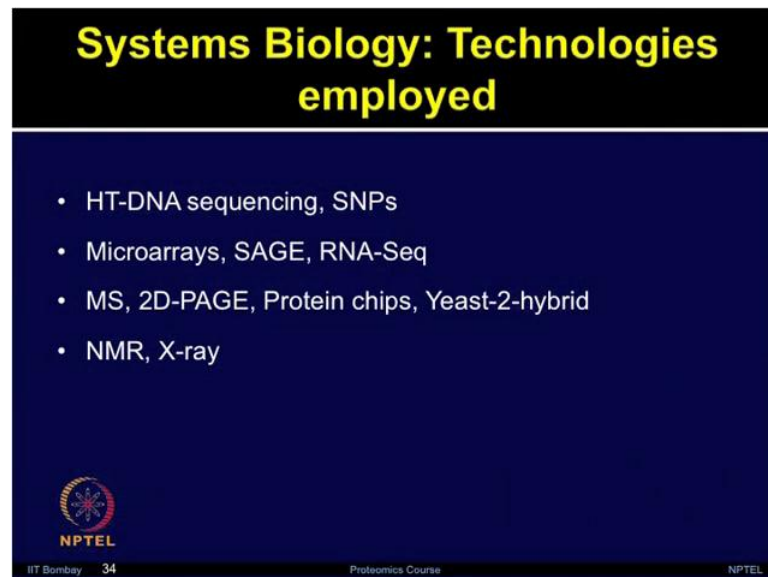
To study the system we need to know about the components and its interactions, the data about the components comes from genomic and proteomic studies, the information about the molecular interactions comes from the interactomic studies, here it is shown that is the systems approach experiment, technology and computational modeling, this triangle is very important which has to be linked with the theory to form a systems triangle.

(Refer Slide Time: 29:57)




So, the different technologies which have been employed to study the systems biology; obviously, you need high throughput data sets which could be derived from microarray platforms, RNA deep sequencing, different configurations of mass spectrometry, different type of structural proteomic tools and protein interaction data sets.

(Refer Slide Time: 30:17)



Systems Biology: Technologies employed

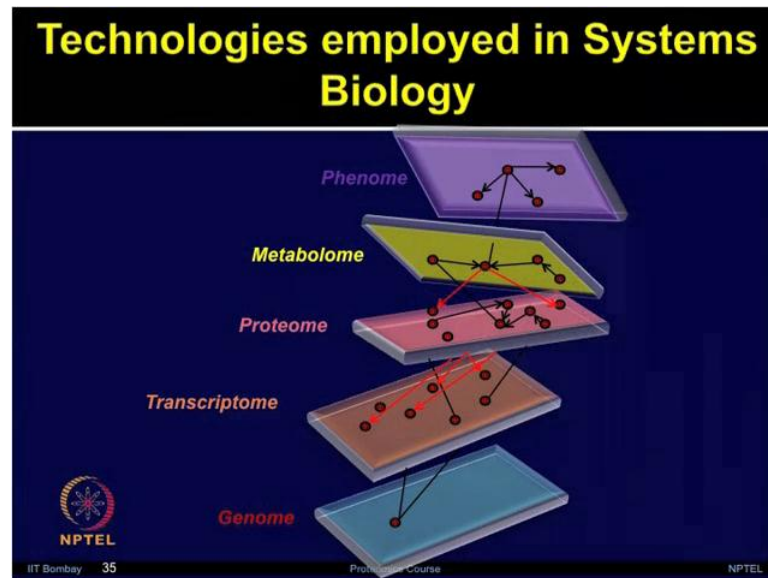
- HT-DNA sequencing, SNPs
- Microarrays, SAGE, RNA-Seq
- MS, 2D-PAGE, Protein chips, Yeast-2-hybrid
- NMR, X-ray


NPTEL

IIT Bombay 34 Proteomics Course NPTEL

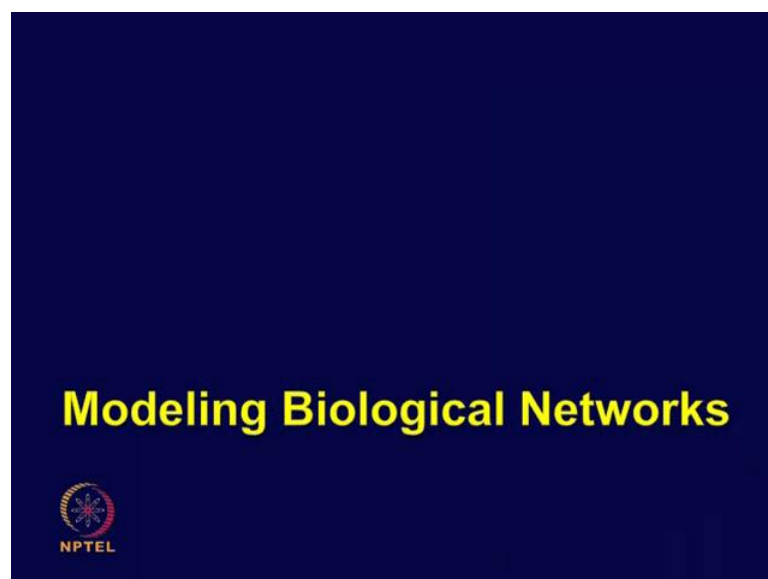
So, some of the technologies which are commonly employed in systems biology can be classified broadly under the following techniques, for genomics the high throughput DNA sequencing, methodologies mutation detection using SNP methods for transcriptomics the transcript measurement can include serial analysis of gene expression Sage, gene chips, Microarrays and RNA sequencing, for proteomics mass spectrometry, two dimensional electrophoresis, protein chips, yeast two hybrids, X-ray NMR is mainly employed for the metabolic analysis the metabolomics.

(Refer Slide Time: 31:06)



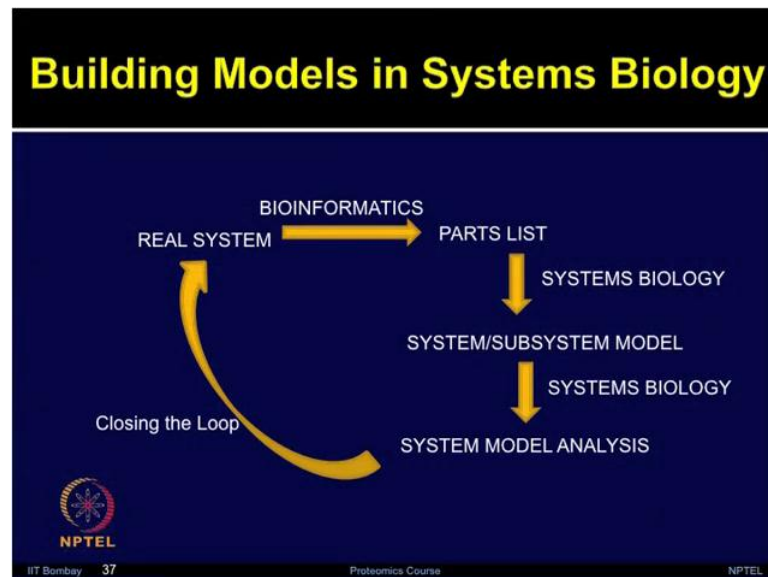
So, as you can see here to generate the systems level information, the systems study requires different type of technologies which could be employed in the biological systems, at genome level by studying different type of technologies using high throughput sequencing, high density arrays, transcriptomics, different type of transcriptome analysis using RNA sequencing and microarrays, proteome we discussed many methodologies, metabolome could be either using NMR or mass spectrometry, and then phenome which is studying about the images by using (()) or NMR methods, so each level of these omic technologies can be useful for studying the systems biology.

(Refer Slide Time: 32:01)



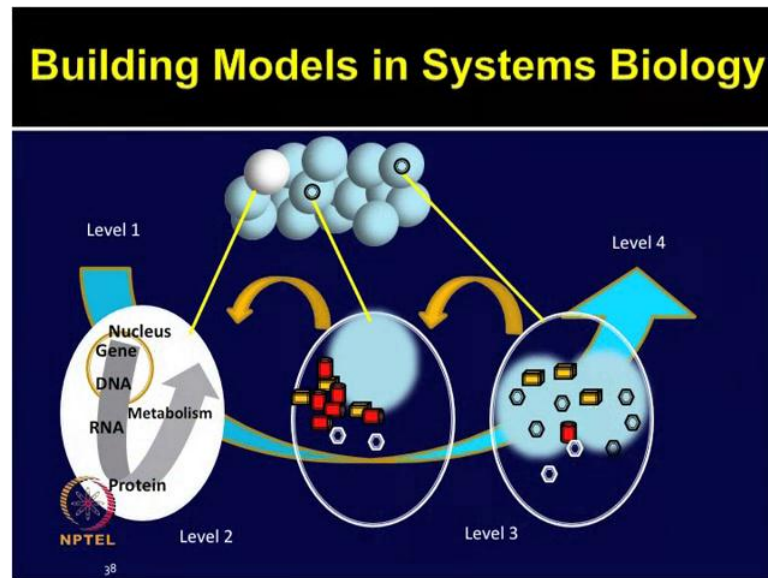
Let us now talk about how to model the biological networks? To build a model in system biology first of all the parts list can be generated by using data sets derived from the systems biology approaches.

(Refer Slide Time: 32:27)



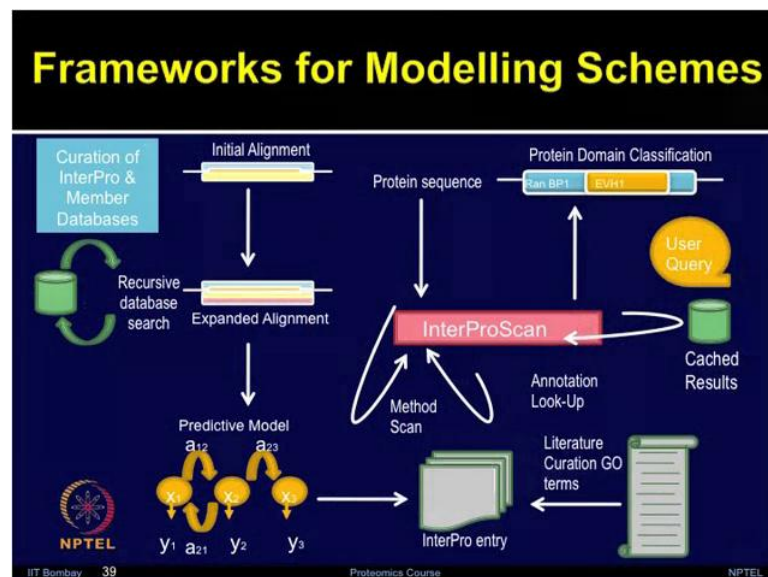
The systems are subsystems model can be generated which can be used for the systems model analysis, now this could be applied for the real systems and by applying the knowledge, using bioinformatic tools it could be again applied back to the original components which could be used to derive some hypothesis and validation of the data sets so it will work like a closed loop.

(Refer Slide Time: 33:00)



To build the models in systems biology information is generated at different levels, level one such as DNA and gene expression, level two the intracellular networks, level three cell cell and transmembrane signals, and level four integrated organ level information.

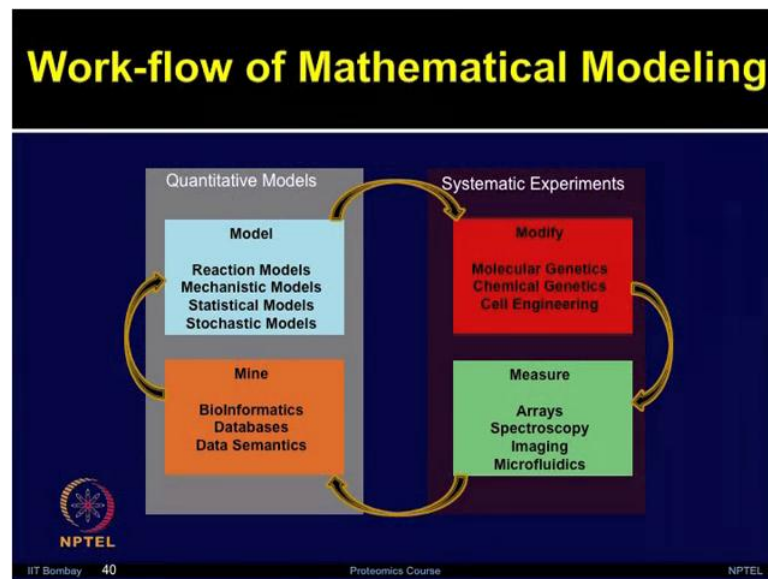
(Refer Slide Time: 33:54)



What is the framework required for the modeling schemes? Different type of deterministic or stochastic models have been proposed, the compartmental variables or individual or functional variables have been studied, the specially homogeneous or especially explicit models are generated which could be applied in the uniform time scale

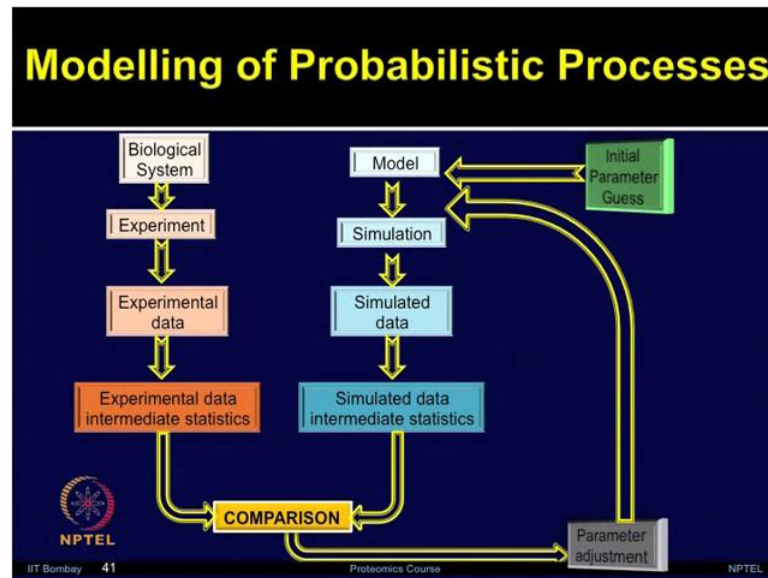
or separated time scales, now this framework could involve single scale entities or cross scale entities, as you can see here this framework requires different level of information in very complex manner, whether it is curation of the databases, how to align these information using bioinformatic tools to generate the predictive models which could be also developed by using the literature curated data sets or experimental data sets; and finally, it could be used to study the systems level properties.

(Refer Slide Time: 34:43)



Let us discuss the workflow of mathematical modeling, a paradigm can be proposed based on modify, model, measure and mine, so systematic experiments different type of molecular genetics, chemical genetics and cell engineering approaches can be used for modifying, and different level of measurements by applying microarrays, spectroscopy imaging and microfluidics, based approaches from proteomics and genomics can be used further for mining which involves bioinformatics, data bases and data semantics, now these data sets could be used to derive the models which could be reaction mechanistic statistical or stochastic models, so starting from systematic experiments to reaching and deriving the quantitative models this work flow can be applied.

(Refer Slide Time: 35:38)



The modeling of probabilistic processes involves let us say you want to study a biological system so some experiments have to be performed, the experimental data sets will be generated from which some (()) can be applied which can be used for the comparison, now different type of models can be generated by using simulations and simulation data sets which can be used for intermediate statistics, and by comparing these two type of information and adjusting the parameters one can study the systems and derive the probabilistic processes.

(Refer Slide Time: 36:26)

ODE and Stochastic models

- Ordinary Differential Equations (ODE): mathematical relation that can be used for modeling biological systems
- Stochastic model: modeling a biological network based on its stoichiometric coefficients, reaction rates and metabolite concentrations

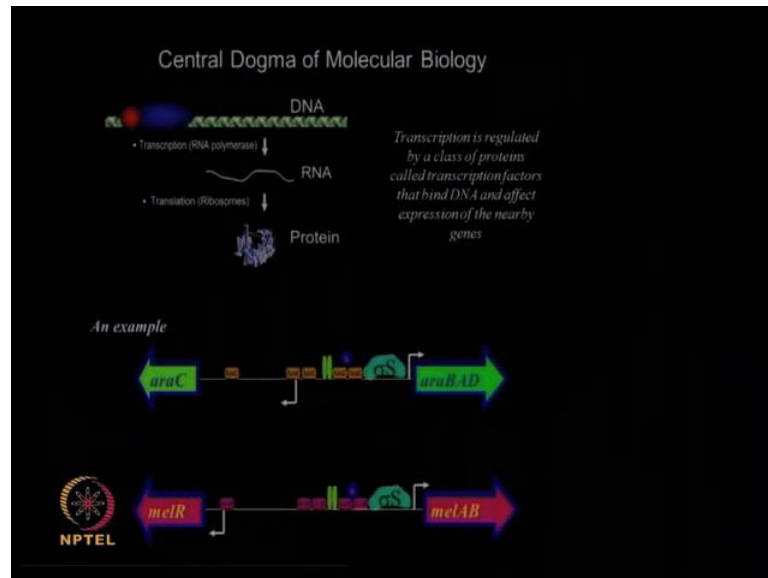
NPTEL
IIT Bombay 42 Proteomics Course NPTEL

So, what is ordinary differential equations and stoichiometric models? The quantitative analysis measures and aims to make models for precise kinetic parameters of a systems network component, it also uses the properties of network connectivity. The ODE is a mathematical relation that can be used for modeling the biological systems, the quantitative models mostly use ordinary differential equations or ODE to link the reactants and products concentration through the reactions reaction rate constants, to develop a computationally efficient and reliable models of the underline g-regulatory networks these ODE models can be used, the stoichiometric model it is modeling a biological network based on stoichiometric coefficients reaction rates and metabolite concentrations.

This is my pleasure to introduce doctor Sarath Chandra Janga from Indiana university and Purdue university Indianapolis, he is in school of informatics and school of medicine. So, as we have been discussing about need for studying proteomics assistant biology till lot of information available at the transcription and translation level, and often this is not good correlation between RNA level and the protein level, so today it will be interesting to talk about systems approaches for a study biological networks from post-transcriptional control towards the drug discovery, so I have invited professor Sarath for having a discussion and a short talk on this topic.

Thank you doctor Srivastava it is my pleasure to be here to talk about some of the work that we have been doing and more generally the principles of regulation and how we can use systems approaches for understanding biological networks more generally, as some of you might be familiar with the use of the concept of networks is increasingly becoming prominent in not just proteomics but also in genomics data and all kinds of high throughput data. So, today what we will be talking about is some basic introduction to the application of networks and biological systems, and how it can be applied to understanding transcription regulation, post-transcription regulation and as well as to the proteomics data, and at large how this can be used to understand the drug discovery how can how it can be applied to the drug discovery concept.

(Refer Slide Time: 39:41)

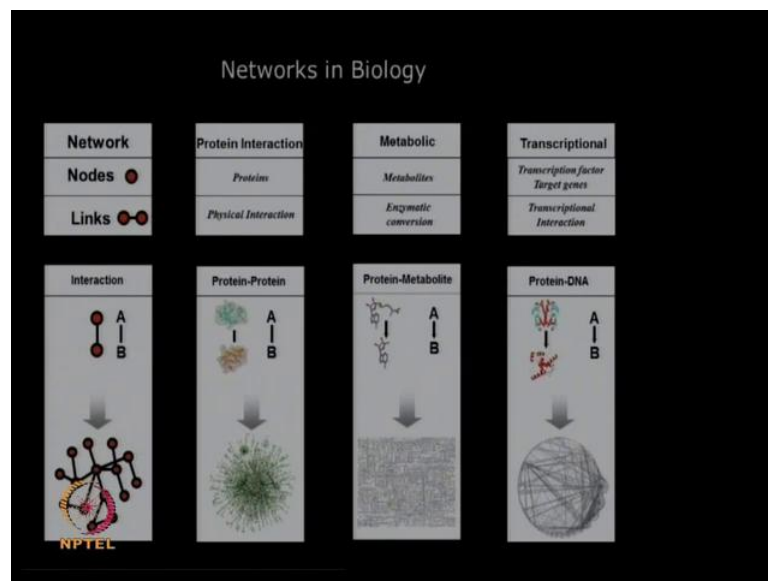


According to the central dogma of molecular biology DNA gives rise to RNA through the process of transcription, and this process facilitated by the binding of the RNA polymerase as well as a number of other transcription factors which bind to the upstream regions of the DNA as we can see and control the expression, and RNA can give rise to protein through the process of translation, and this happens through the process of translation with the help of ribosomes, now in this process the proteins which I have produced some of them can be classified as transcription factors which bind to the DNA, and some others classified as RNA binding process which can bind to the RNA and control the expression at the post-transcriptional level as oppose to at the transcriptional level where transcription factors bind to the DNA.

Now, as an example let see that case of arc transcription factor in a bacterial genome such as e coli, this particular transcriptional factor binds to the upstream reasons of ara b a d operon which encodes for the enzyme and the transporter responsible for uptake of arabinos from the environment, now the transcription factor ara c not only binds to the upstream of ara b a d, but it can also binds to itself and control expression. As you can see from the small orange boxes which are shown as representation for the binding sights, now what they suggest is transcription factor can out regulate bind and regulate it is own expression or it can also bind to other gens controlling their expression, there are also cases there are many cases actually where transcription factor multiple transcription factors bind to the upstream regions.

As you can see in the case in this case represented with the orange box as well as the blue box so blue circle where other transcription factors bind, now in addition to this binding of transcription factors as I mentioned earlier polymerase RNA polymerase also binds shown in with the with the green box a green circle green oval box out there so that they can control the expression, now there are other examples I also shown in this figure with mel r regulator also doing something similar.

(Refer Slide Time: 41:48)



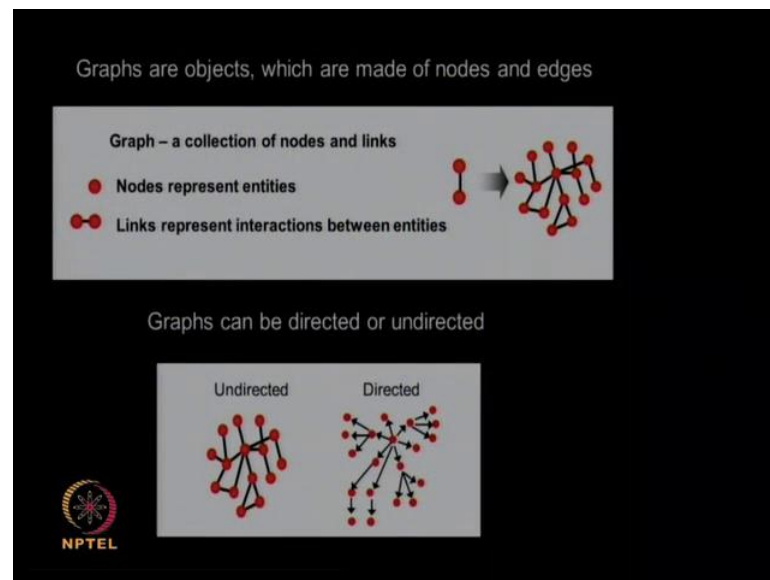
Now, this is what we just discussed is an idea of how regulation happens from a biological view point, now an increasing thing increasing amount of literature now supports the ideas of networks in biology, so what exactly our networks? Networks simply represent are represented as nodes and links or edges, this nodes can be biological entities and the links or edges or actually the association between these entities, now there are number of ways you can talk about the nodes or the entities, so one common form of representation are protein interaction networks, where the proteins form the nodes and the physical interaction between this proteins forms the edge as you can see in this in the figure below.

The you can have a representation of this networks in a in a fashion that is shown in this figure below, now an alternate kind of network which is also studied in the literature over the last 10 years or so are metabolic networks, in metabolic networks the metabolites form the nodes and the conversation of one metabolite to other forms the

edge in this case, now as you can imagine the conversion of one metabolite to the other is actually facilitated by the enzyme, so the particular protein enzyme converts a metabolite a to b, and you look at on a global scale and when you looking at the conversion number of metabolites one to the other and sometime one metabolite can give rise to more than one set of a metabolites, such complex set of associations can be called as a metabolic network.

Now, the third kind of networks which I will be elaborating in more detail in the next slides a transcriptional networks, in transcriptional networks transcription factors form one set of nodes and the target genes form other set of nodes, so as you can imagine what here actually looking at in this case from a biological viewpoint is the interaction of the transcription factor with the DNA and controlling of the expression of the downstream genes, but in the context of the networks what we are showing here what we are showing is the transcription factor and the target gene or operon whose expression is control, again in this case you can see that the a protein a which is transcription factor controls b, but it may or may not b that b is transcription factor and it also controls a, so that might be a case to case specific and may or may not be having reciprocal interaction.

(Refer Slide Time: 45:05)

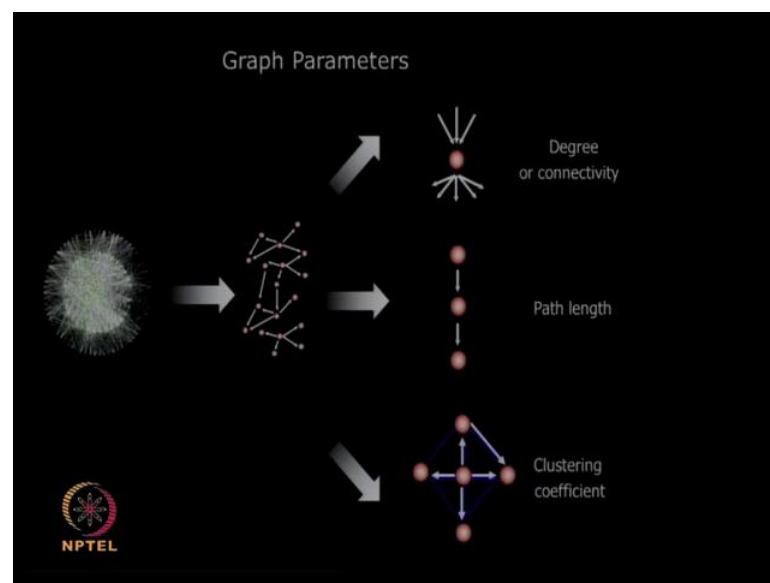


As we just discussed these networks are actually the concept of network has been borrowed from physics and computer science where often this kind of networks are refold to as graphs, and graphs are objects which are on collection of nodes and entities

the nodes are representing the entities it could be this these entity could be genes proteins small molecules cells organs or at any level you can represent these entities, the interactions are association between them or the links, now as I am just mentioned there are different kinds of networks there protein-protein interaction networks, metabolic networks, transcriptional networks.

In the case of protein-protein interaction networks what we are looking at often is no directionality in such interactions and these are called as undirected networks; however, there are also directed networks such as transcription networks or metabolic networks, in these cases there is a flow of information I e where a controls b which should mean a is controlling a is regulating b so therefore, there is a directionality, and these are a often commonly studied as regulatory networks.

(Refer Slide Time: 45:27)

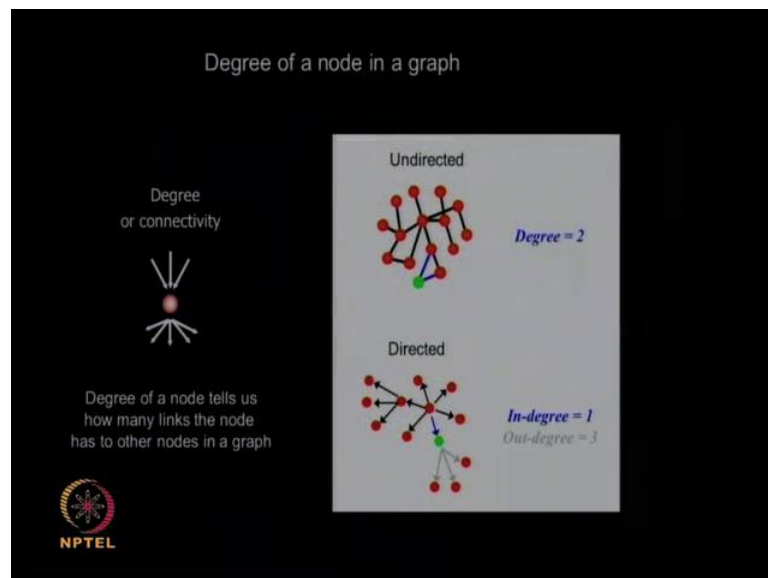


And we will be talking in more detail in the next slides; however, before we get into the more specific observation about the properties of this networks one set of common properties which are studied when you are looking at biological networks are degree, path length and clustering coefficient.

Now, often when you look into a network as such you do not have a clear understand of the properties of the different nodes, but when you look into the specific aspect such as

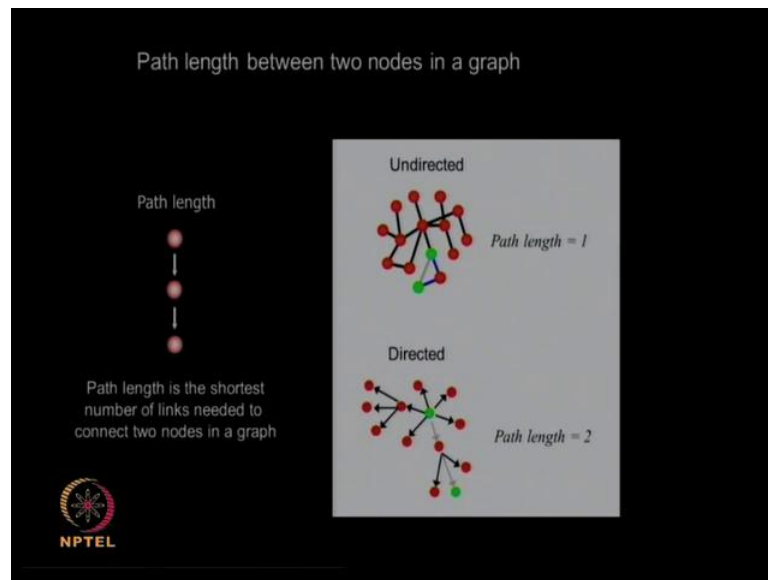
in this case shown in this case as degree what it tells you is how many connections of particular gene protein or node has in your network? So, what we can say from the first example on the top is the degree of the node is 8; that means, it is connected to 8 other proteins, and in the second property is path length, what it is showing in this case if you is that the number of edges that you need to travel from one node to the other, so if I ask you what is the path length between that first and the bottom node in this figure you would say the path length is equal to two, the third kind of property which often studied is the clustering coefficient; clustering coefficient tells how often the neighbors of a given node are connected to what you would see in a completely connected graph.

(Refer Slide Time: 46:48)



Let us look at more detailed examples, for instance if you are studying the degree of a node in the case of undirected network such as in the examples shown in the top the fluorescence node that shown fluorescence color has a degree equal to two, on the other hand a directed node example shown at the bottom has a degree equal to four, because it is connected to four other nodes; however, what you can also say is there is in degree and out degree, and in degree is the number of incoming connections of a particular node. So, the green fluorescence nodes here has in degree of one, it also has an out degree equal to three because it is directing three other nodes shown in red color out there, so it is out degree equal to three.

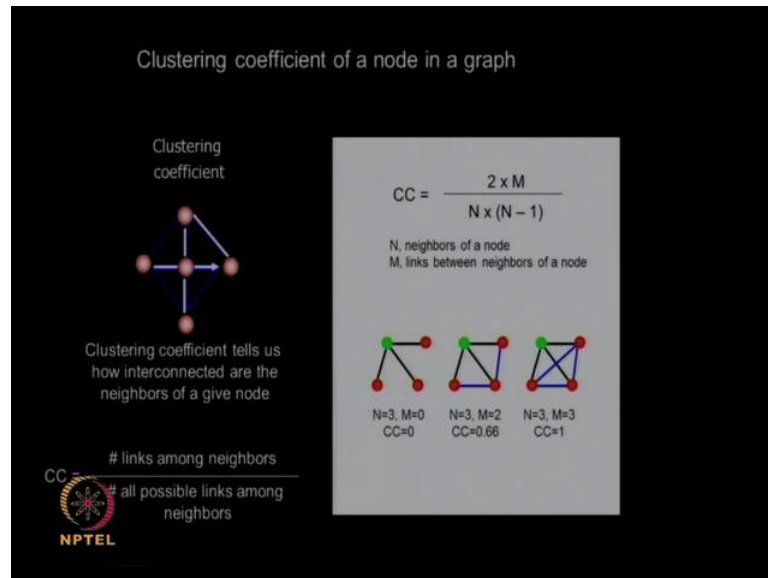
(Refer Slide Time: 47:28)



Now, you can also extend this idea of undirected and directed graphs and ask what is the path length of a node, now as I mentioned the path length is referred to as the number of edges that one need to travel between two different nodes that you are interested, on in the top a network that you are seeing the path length between the two green or florescence nodes is equal to two as well as equal to one, because the path that you can take can be different then the shortest path that you looking at; however, almost often unless they were specified when you are talking about the path length between two nodes it is the shortest path length, so the two florescence nodes have a path length equals to one; however, if you if you are ask what are all the path lengths you would say they it has two different paths one with a path length of one the other with a path length of two.

In the undirected networks your definition of path length essentially does not change, so in the example that you see at the bottom the path length between two florescence nodes is equal to two.

(Refer Slide Time: 48:33)

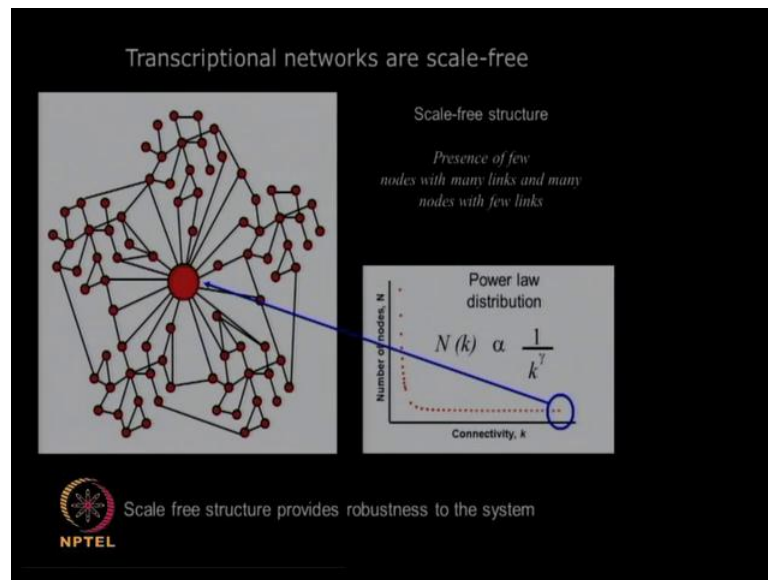


The other property that I was referring to previously is the clustering coefficient of node, and clustering coefficient refers to the number of the connection between the neighbors of a given node of interest to what you would see in a completely connected graph.

Now, let us look an example in this figure that you see there is the first example, the florescence node the green node has three connections three red dots are connected to it; however, if you ask the number of connection between the red dots it is 0, there are 0 connections between the red dots, but if they were fully connected you would see that they will have three edges between them, so the clustering coefficient of the florescence node right now is 0 upon 3, let us look at this second troy network in the second troy networks the florescence node has a clustering coefficient of 2 upon 3, the third case the clustering coefficient of the florescence node is 3 upon 3 which is completely connected, so the clustering coefficient is equal to 1.

Now, more generally formula can be brought up and it can be written as if there are n number of interactions between the neighbors of a node of interest and there are n number of neighbors of a given node of interest then it can be written as m upon n into n minus 1 by 2, so that would be defined as a clustering coefficient of that particular node, and on average the clustering coefficient of a node on a whole network scale it gives you an essence of modularity of the network, the higher the average clustering coefficient the more likely is the network cluster can be decomposed into specific a modules.

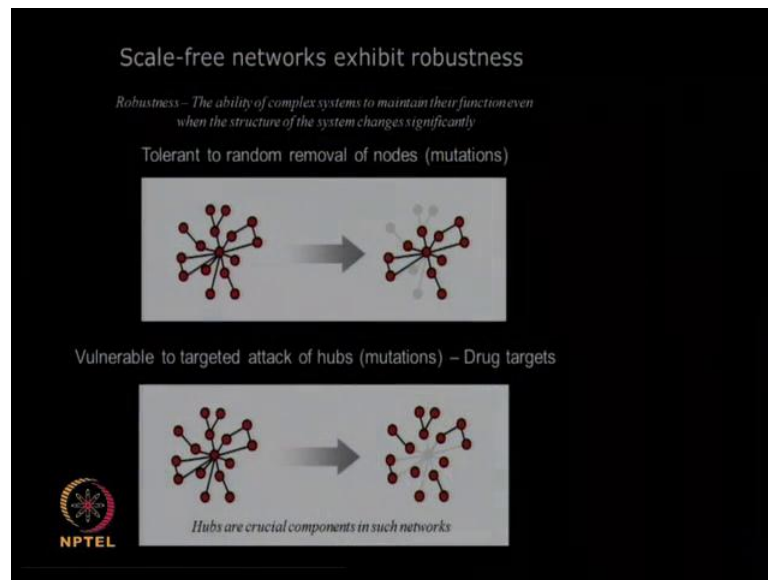
(Refer Slide Time: 50:24)



Another property that is of great interest in understanding biological networks is a scale-free structure, and while lot of biological networks are documented and shown to be scale-free transcriptional networks are also documented to be scale-free structures, so what exactly our scale-free networks? Scale-free networks are correspond to the structure of a network where there are few nodes which are highly connected, for instance in the figure to the left in the network figure that you see to the left there is a big red dot big red node which is highly connected so but there are not many such highly connected nodes, and there are many nodes which are very poorly connected.

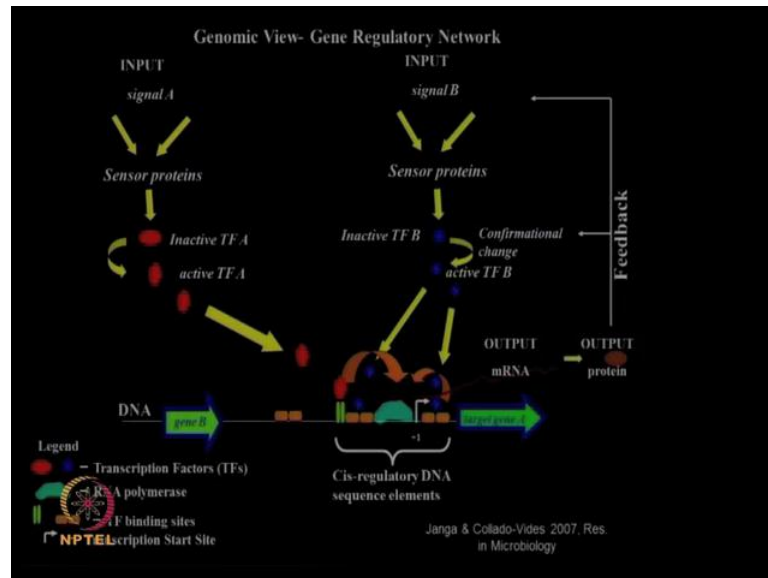
So, in other words a scale-free structure refers to a network structure where there are few nodes which are highly connected and most nodes are poorly connected, or more mathematically if you plot the connectivity of a node versus the number of nodes with a given connectivity we should see a power-law distribution, otherwise if you plot the log-log plot of the connectivity versus the number of nodes with a given connectivity you should see a negative slope of gamma as shown in this figure, where gamma lies between 2 to 3, that is when you can call the structure to be scale-free and the and the distribution to be a power-law distribution.

(Refer Slide Time: 52:01)



Now, what is so special about this scale-free structure? Scale-free structures have been postulated to provide robustness to the biological system. Now, what exactly is robustness? So, robustness is the ability of a complex system a complex system such as a biological system to maintain it is function even when the structure of the system changes significantly, now let us look at an example so in the network figure that you see if you randomly perturbed any of these nodes you are likely to effect a small fraction of the network; however, if you target the highly connected node that is the central node which is highly connected you are going to disrupt a major fraction of this network suggesting that these highly connected nodes can be vulnerable to be the drug targets, so if you are trying to inhibit the growth of a pathogen you are likely to target this highly connected nodes because you are more likely to be able to crumble the biological system of the pathogen, so and this has been increasingly gaining attention as a method of targeting drugs to this kind of this class of proteins.

(Refer Slide Time: 52:47)



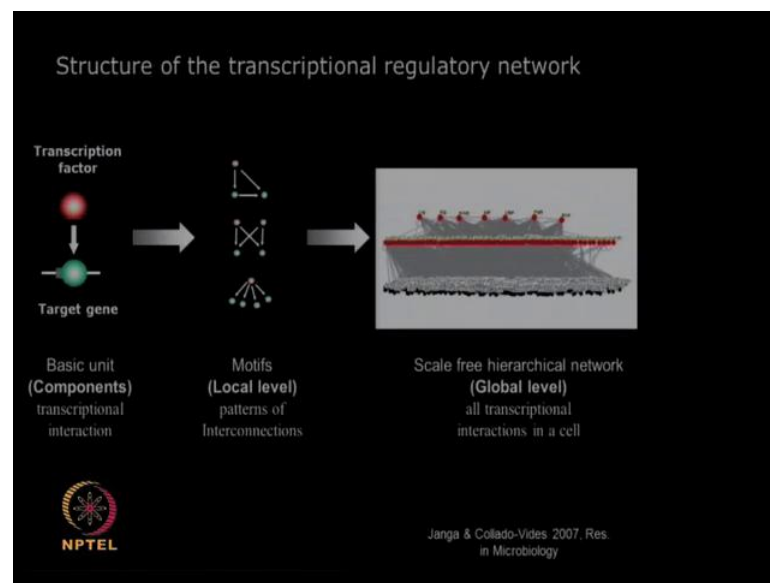
So, as mentioned earlier we have been talking about regulation of a single transcription factor, but in the context of network regulations is much more complex and what we are referring to is a combinatorial regulation by many different transcription factors, let us look at a specific scenario, so the slide shown here shows a typical regulatory system in a bacterial organism, what you usually have is a set of signals which are sense by the cell and these signals are sensed by sensor proteins, these sensor proteins could be transcription transporters, or this could also be (()) kinesis.

And once these sensor protein sense the signals from the exterior or even sometimes interior of the cell they can cascade the information to transcription factors, the transcription factors upon receiving the signals can change from active to inactive or inactive to active state, and when this happens because of multiple sensor proteins these transcription factors can change the confirmation and bind to the upstream regions, and shown at the bottom is a stretch of DNA where this transcription factors can bind in combinatorial fashion often and control the expression of the target gene or operon, as a rule of thumb if transcription factors bind to the upstream regions in the upstream of the transcription start site shown as plus 1 that is where the transcription actually starts, you often or stimulating the polymerase and enhancing the expression.

However, when you bind to the downstream of transcription start site you typically repress the expression of the target gene thereby blocking the transcription by the

polymerase shown in the oval shaped polymerase symbol in green, so based on these principles and together with the interplay with the transcription factors and the polymerase your transcript is produced, and once transcript is produced you can have mRNA and protein levels regulation which is not what will be talking immediately now, but all these levels together contribute to provide feedback and this is typically system simple regulatory system that you encounter in bacterial organisms, but more complex systems more complex eukaryotic gene regulation is much more complex and beyond the scope of our current discussion.

(Refer Slide Time: 55:09)

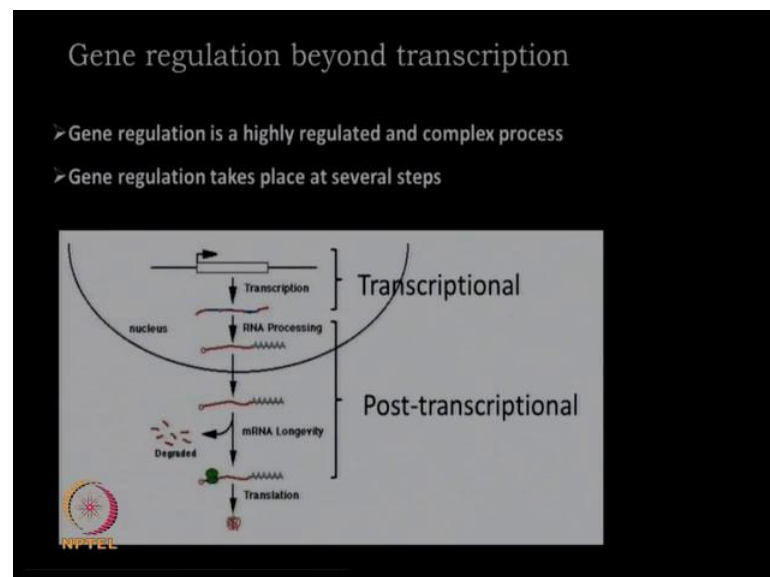


As discussed in the previous slides the basic unit of regulation is a transcription factor and a target gene whose expression is being controlled, now on a different scale if you increase if you put together all the set of regulatory events between transcription factors and the target genes are operons you construct a global transcription regulatory network, and as I mentioned earlier this network is a scale-free structure scale-free network, but in addition to this it is also hierarchical structure wherein what we are actually referring to in a hierarchical structure is there are set of transcription factors which are able to regulate a large number of genes, and there are set of genes other transcription factors which are also controlled by this global transcription factors shown at the top of this network structure, and both the top layer and the second layer all of them together regulate the set of genes which are not essentially encoding for the protein coding which are not essentially encoding for the transcription factors.

So, in a way there are transcription factors which are at the top of the system, there are transcription factors which are controlled by this top layer and there are subsequent layers, and the number of layers such a hierarchical structure depends on the complexity of the system, now in between the top and the bottom layer there are set of substructures or sub graph within the regulatory network which we call as motifs. Motifs are the set of sub graphs which occur more often than expected by chance, and there were three kinds of regulatory motifs that are identified in regulatory networks.

One is the feed forward loop where there is there are two transcription factors the first transcription factor regulates the other two genes, the second transcription factor regulates the target gene. The second kind of motif is multiple input modules where there are two different transcription factors both of them regulate two different target genes. The third is a single input module where a single transcription factor regulates a set of target genes, now each of this set of regulatory motifs have been shown a specific function and which would be beyond the scope of our current discussion.

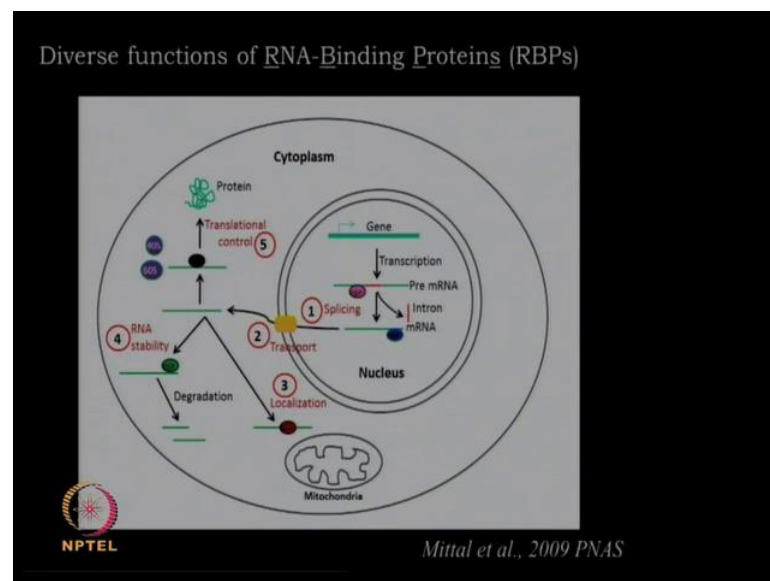
(Refer Slide Time: 57:26)



Now, although the idea of regulation of gene expression the level of transcription has been documented for several years and we have extensive understanding very little is known about the regulation of gene expression beyond transcription, and it has only been recently being appreciated about the role of regulation at the post-transcriptional level,

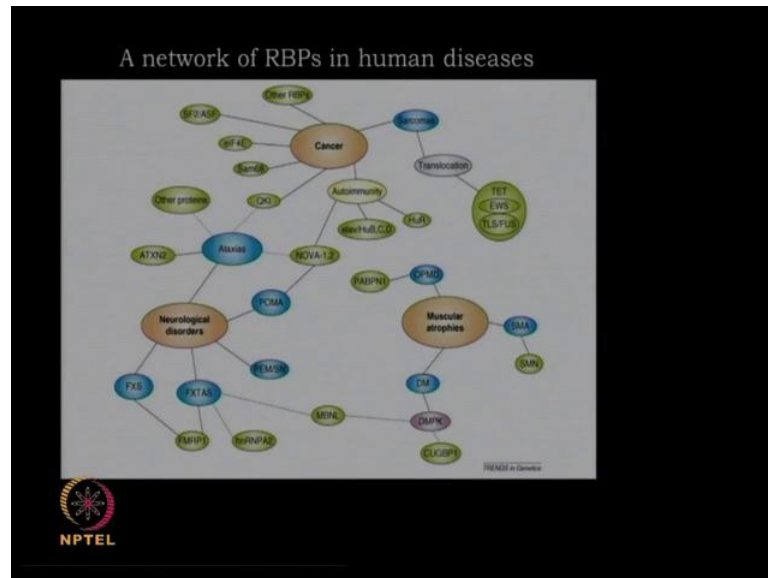
now most of this evidence for the reason why post transcription regulation is becoming important is coming from the lack of correlation between mRNA and protein pools in model systems, now there is now enough evidence to suggest that this post-transcriptional processes are actually controlled by a class of proteins called RNA binding proteins, among non-protein coding components such as micro RNA's non-coding RNA's, so RNA binding proteins are now known to be involved in controlling the RNA processing, RNA longevity as well as in the translation.

(Refer Slide Time: 58:26)



Now, in particular as soon as the gene is transcribed and pre mRNA is produced splicing associated RNA binding proteins bind to the pre mRNA and convert into mature mRNA by splicing of the introns, now the produced RNA not necessarily only mRNA is needed to be exported from the nucleus into the cytoplasm, and this is carried out by a class of RNA binding proteins which can be termed as transport RNA binding proteins shown with number two in the figure, RNA binding proteins have also been implicated in the specific sub cellular localization of these transcripts, RNA binding proteins are documented also in controlling the stability of the transcripts thereby promoting or degrading the expression of these transcripts, as expected RNA binding proteins a number of them are associated with the ribosomal proteins to control the regulation of expression at the translational level.

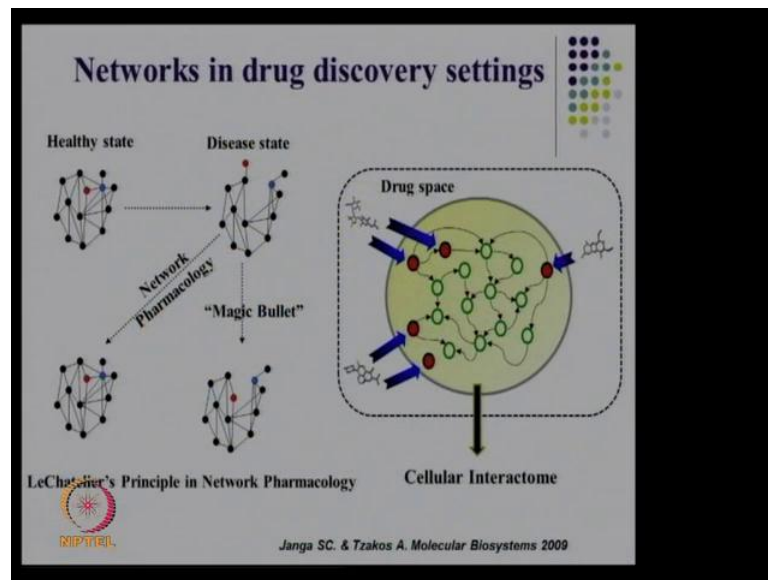
(Refer Slide Time: 59:24)



Now, other aspect of regulation understanding regulation at the post-transcriptional level is that number of RNA binding proteins are involved in human diseases major class of human diseases such as cancer, muscular atrophies and neurological disorders, in this network diagram shown here the major class of diseases are shown in orange while the subtypes of diseases which are which can be sub classified or shown in blue, and the specific RNA binding proteins which have been documented are implicated in these disorders are shown in green.

Now, let us take it a specific example of as a muscular atrophy called myotonic dystrophy, in this particular kind of disorder a CUG repeat binding protein called CUG B P 1 binds to the three prime un translated region of a D M protein kinase, and because of the sequestration of this CUG repeat binding protein onto the trinucleotide repeat expansion in the three prime un translated regions this particular disease phenotype is observed, another example of misregulation of an RNA binding protein happens in OPMD which is another kind of muscular atrophy, in this particular kind of disease there is a GCG repeat expansion in the axon one of an RNA binding protein which is a poly a binding protein called pab P N 1.

(Refer Slide Time: 01:01:15)



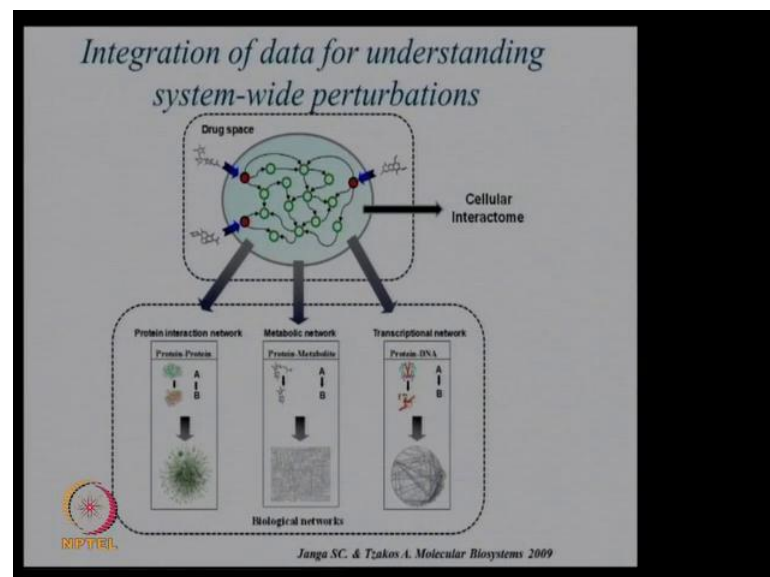
Another example we can observe which is which is heavily documented in the literature is a brain specific splicing factor called nova, whose misexpression is known to cause a disease called poma which is a subtype of neurological diseases, so what I am trying to (()) here is that if there is a change in expression of either RNA binding protein, or any of its targets it can be associated to a disease phenotype, and all the studies basically such is that it is not just the effect of a single gene or protein it is rather a combination of different set of genes and proteins which contributes to a disease phenotype, now while this observation is not very new while we knew that this is common for a number of complex diseases, what we have still been not able to achieve is they able to cure diseases for these complex diseases.

Now, let me introduce to you the traditional notion of how drug discovery is usually happening in most places, let us represent the healthy state of an individual with a network of interactions shown in this figure on to the left, now at disease state could be studied as a perturbation such a network where some of these nodes are actually not properly connected compared to the healthy state, now according to the idea of Paul Ehrlich and others the magic bullet approach suggest that the conversion of disease state to the healthy state should involve one or most likely one particular drug which is non-promiscuous and specific to a particular drug target, so that you have minimal off target effects.

Now, often such magic bullet approach can only yield only a semi-recovery to the to the from the disease state, now what network pharmacology or network medicine approaches are trying to arrive at is use a combination of perhaps promiscuous drugs, but which do not cause negative side effects, which do not cause side effects with a lethal and can still convert the disease state into healthy state as close as it is to the original one, now how would you achieve such an approach? To understand the whole this particular idea let us look at a network representation of how the different entities in the cell are interacting, in the figure to the right you can see that a number of drugs each of them can be perturbing different nodes.

Now, all of these nodes are actually interconnected to each other because we are looking into the cellular contacts and there are protein-protein interactions; there are metabolic interactions; there are also regulatory interactions perturbation of one cannot be seen in isolation it has to be seen in the context of other perturbations, now a combination of these perturbations is going to yield phenotype which we hope can be treating the complex disease that is the concept behind this idea of network pharmacology.

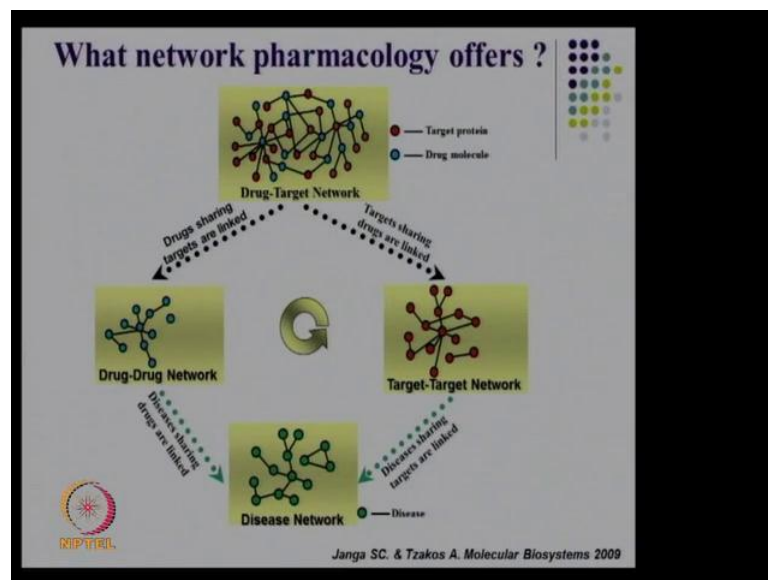
(Refer Slide Time: 01:03:44)



Now, how do we achieve such bigger goal? So, usually when you have such kind of complex problem complex phenotype you have to put together data such as knowledge on the current metabolic network in the human genome, knowledge on the transcription network, knowledge on the protein-protein interaction network, and knowledge on the

post-transcription network, and together with a current knowledge of the drugs and the targets and the target pathways one can start looking at how these perturbations can be studied in the context of specific diseases, and what particular drugs can be used to identify potential new therapies for existing diseases.

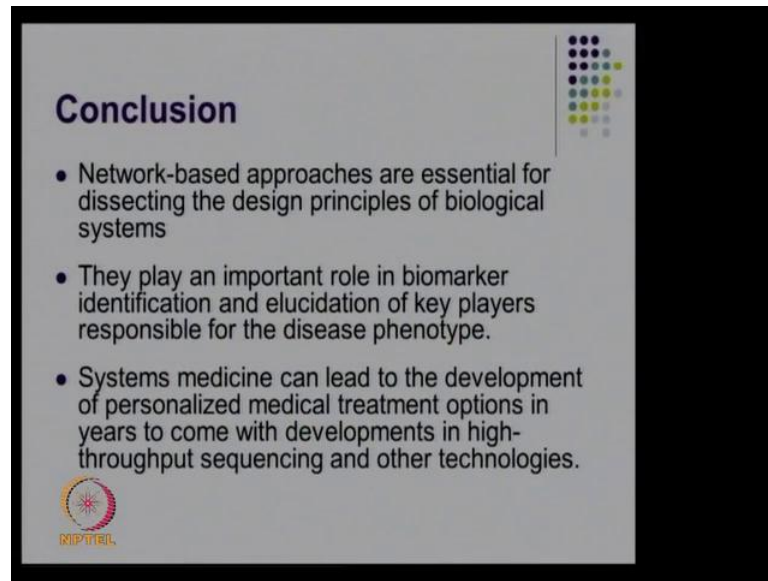
(Refer Slide Time: 01:04:24)



An alternative set of approaches which are being used in the context of drug discovery is that if you have a target a drug target network for all the approved drugs in the literature one can start understanding what are the drugs which are sharing the targets, can we use the drugs which share the targets as alternatives to existing drugs if there is any resistance acquired for a particular drug can you compliment the current drug with another drug which is having the same set of targets, or one can start studying the set of drug drug relations if there drugs are sharing the targets can we start studying what are the profiles of the two true drugs which are linked, are they similar in the structure are they similar in the final phenotypes or what are the common principles of these drugs which are connected to each other.


Likewise one can also study disease disease associations by linking any pair of drugs which are working which are used for the same disease. Likewise one can study target target network to construct disease disease association network. So, this is these are some of the ideas which where the field is moving to understand or to even repurpose existing drugs for novel therapies.

(Refer Slide Time: 01:05:37)



Conclusion

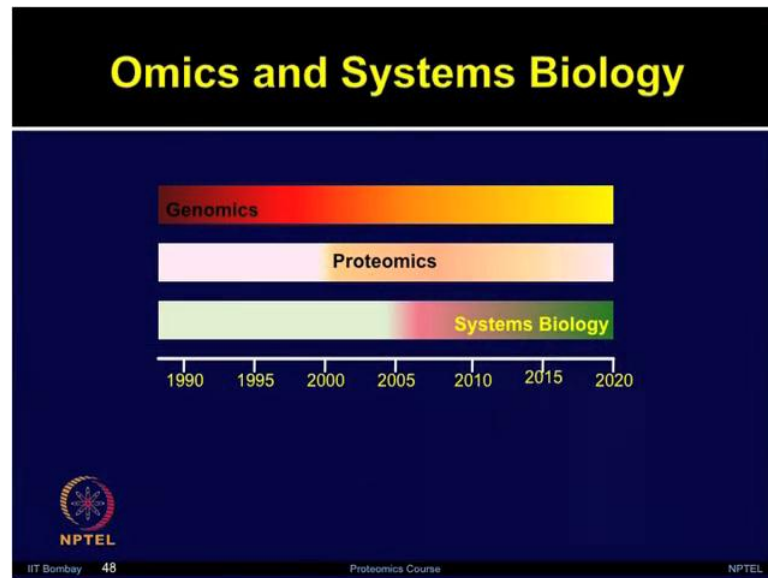
- Network-based approaches are essential for dissecting the design principles of biological systems
- They play an important role in biomarker identification and elucidation of key players responsible for the disease phenotype.
- Systems medicine can lead to the development of personalized medical treatment options in years to come with developments in high-throughput sequencing and other technologies.



So, to conclude what we have tried to cover in the past set of slides is that the network-based approaches are essential and a powerful paradigm for dissecting the design principles of biological systems, they play an important role in biomarker identification and even in the elucidation of key players responsible for the disease phenotype, systems medicine can lead to the development of personalized medical treatment options in years to come with developments in high throughput sequencing and other technologies which can yield a lot of data in a very short time, so that clinical relevance can be achieved for based on these kind of techniques application these network based approaches in the context of clinical settings.

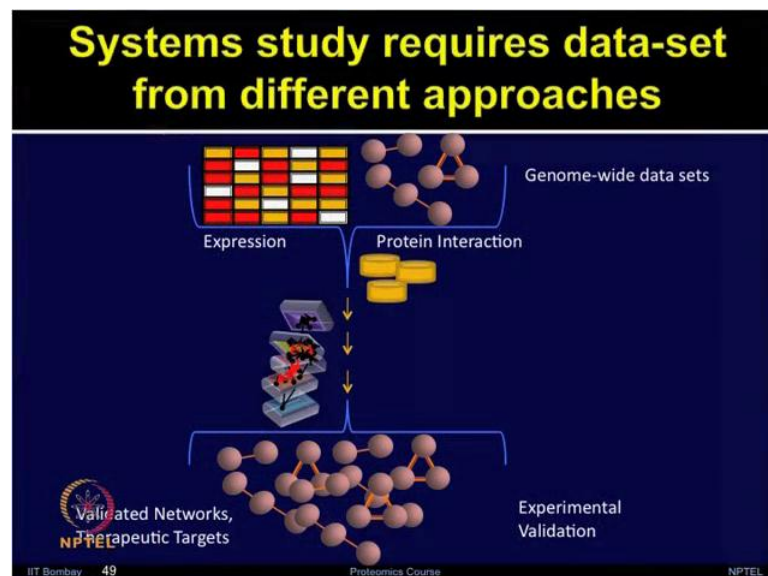
Thank you very much Sarath for giving very nice talk and giving some of the basic concepts as well as illustrating how systems level network studies can be employed for illustrating various types of problems including in the drug discovery as well as in pharmacology and it could standard for even by (()) discovery and many other applications. So thank you very much. Thank you.

(Refer Slide Time: 01:07:06)



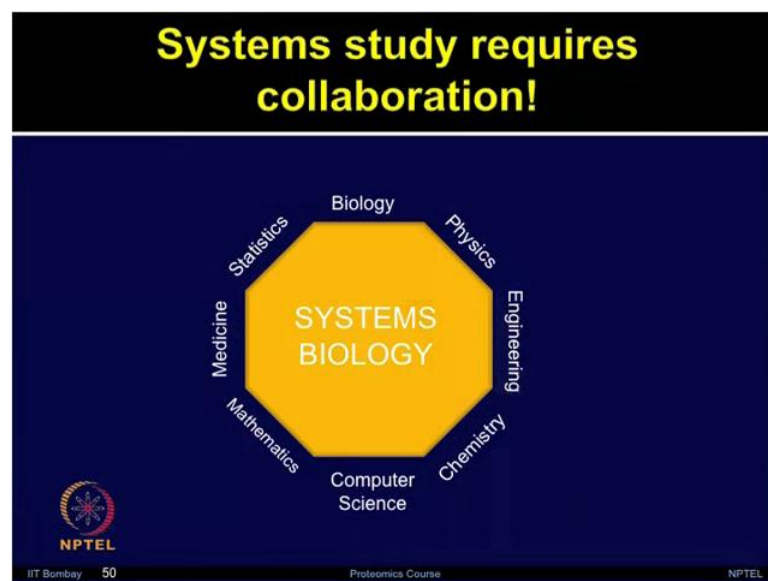
Now, let us try to integrate the omics approaches with systems biology, so genome sequencing projects in genomics era from 1990s to 2000 accelerated the of omics research, then from 2000 onwards proteomics field also got accelerated and new methodologies new tools came into the place for studying the proteome, and the data derived from genomics, transcriptomics, proteomics, metabolomics and other omics approaches have now brought the integration of the data sets in the systems biology field.

(Refer Slide Time: 01:07:44)



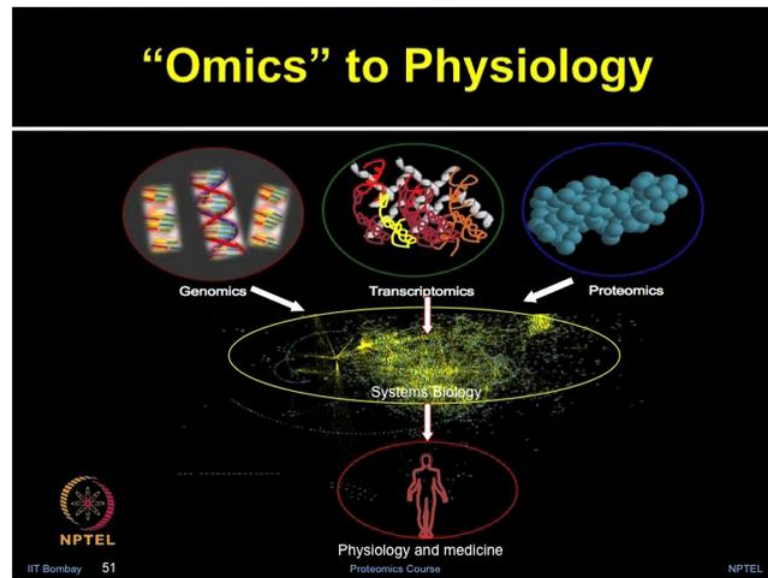
The systems study requires obtaining data sets from different approaches and analyzing them for example, as shown in the slide the genome-wide data sets can be derived at the genome level and looking at the expression of the different transcripts or at the proteome level looking at different type of protein interactions, these data sets can be stored in the clinical databases and also it can be mined from the literature, literature manual curation then integration of the orthogonal data sets further can be used for validating the networks and deriving identifying therapeutic targets, further it can be used for experimental validation.

(Refer Slide Time: 01:08:26)



Studying systems cannot be done in isolation in individual labs, it requires different expertise and collaborations from scientists from different disciplines of biology, physics, engineering, chemistry, computer science, mathematics, medicine, statistics and many more.

(Refer Slide Time: 01:08:48)



See eventual aim of this goal of this current systems biology field is to employ the omics level information obtained from different levels from genome transcriptome and proteome derive that information at the systems level, integrate quantitative models and then propose and use it for the understanding the physiology and apply that in medicine, so this omics to physiology this flow can be well maintained by employing systems biology tools.

(Refer Slide Time: 01:09:28)

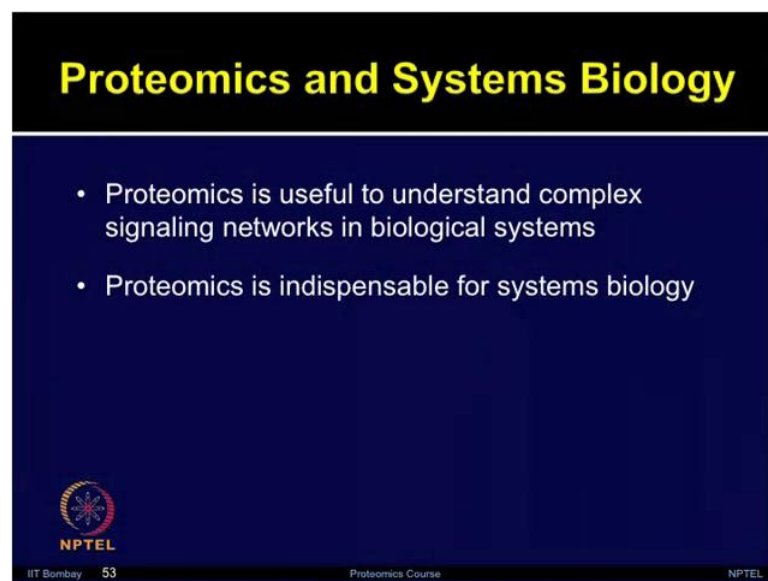
The slide, titled "Challenges: Systems Biology", lists three main challenges in the field. The challenges are:

- Systems biology is extremely challenging
- Understanding dynamics of biological networks requires modeling, simulation and understanding of biology
- Requires mathematical & statistical approaches

The slide is part of an NPTEL presentation from IIT Bombay, slide 52, for the Proteomics Course.


What are the challenges of systems biology? Systems biology is extremely challenging the emphasize to understand a system, understanding dynamics of even simplest biological networks not only requires only the understanding of biology, but also it is modeling and simulation, the disintegrative study can be used for studying from cells to proteins to gene, or integrative study could be used for putting these places back together again, and then understanding and doing the prediction and control of functional biological processes, so all of these are very challenging but currently being addressed by applying various systems level tools.

(Refer Slide Time: 01:10:24)



Proteomics and Systems Biology

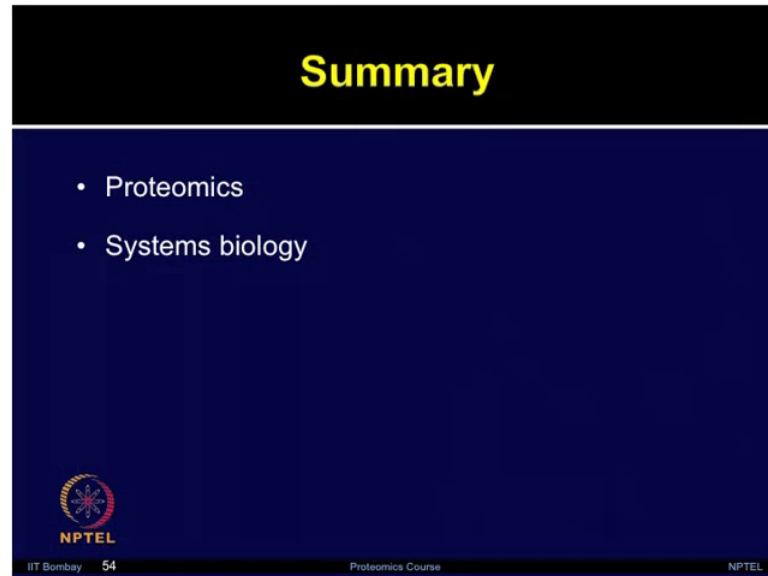
- Proteomics is useful to understand complex signaling networks in biological systems
- Proteomics is indispensable for systems biology


NPTEL

IT Bombay 53 Proteomics Course NPTEL

So, how proteomics and systems biology are integrated? Proteomics as we have studied it is useful to understand the complex signaling networks in biological systems, it is very indispensable tool for systems biology, the global analysis of proteome is important; however, there are many limitations in each experiment only thousands of proteins can be studied therefore, new approaches and systems level investigation and predictions are required, the system investigation is required to study the complex dynamic structure interaction with the biological systems whether it is at cellular level or at the organism level and ultimately it is responsible for their function and behavior.

(Refer Slide Time: 01:11:22)



The slide features a black header with the word "Summary" in yellow. Below the header, on a dark blue background, are two bullet points: "• Proteomics" and "• Systems biology". At the bottom left is the NPTEL logo, and at the bottom right is the text "NPTEL". The footer contains "IIT Bombay 54" on the left, "Proteomics Course" in the center, and "NPTEL" on the right.

So, in summary today we discussed that how omics era the technological advancement in genomics proteomics and metabolomics have generated large scale data sets in all the aspects of biology, these large data sets has motivated the computational biologist and systems approaches with objective of understanding the biological system as a whole, while proteomics continues to generate the quality data at proteome level, so systems biology approach characterizes and predicts these dynamic properties of biological networks. Now, the next module we will focus in more detail different type of proteomic technologies. Thank you.