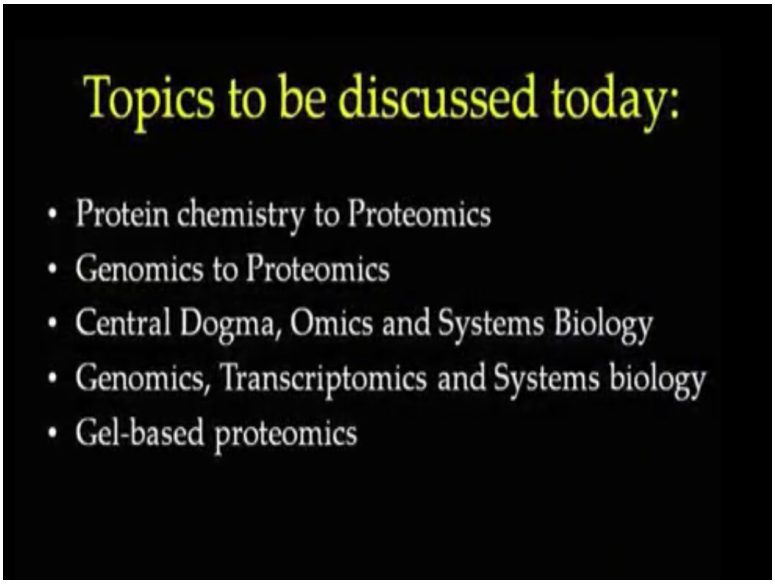
Introduction to Proteomics

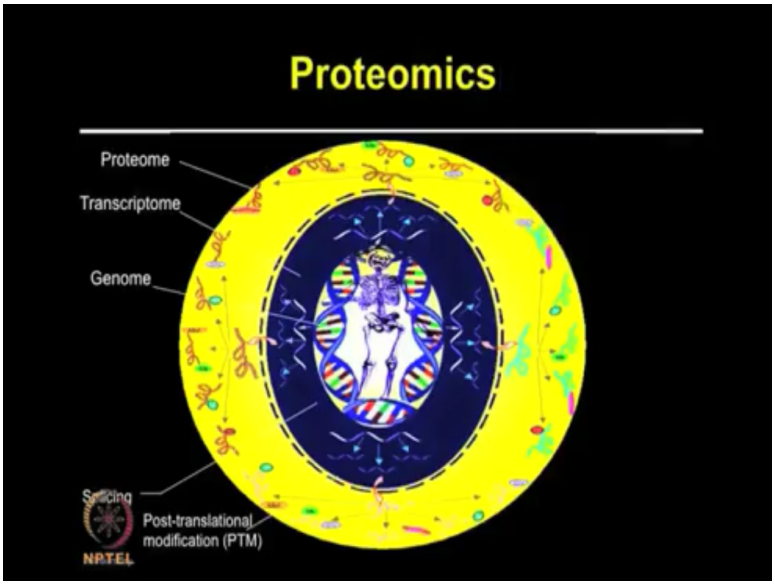
Dr. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology – Bombay

Lecture - 04
Introduction to Proteomics

(Refer Slide Time: 00:15)



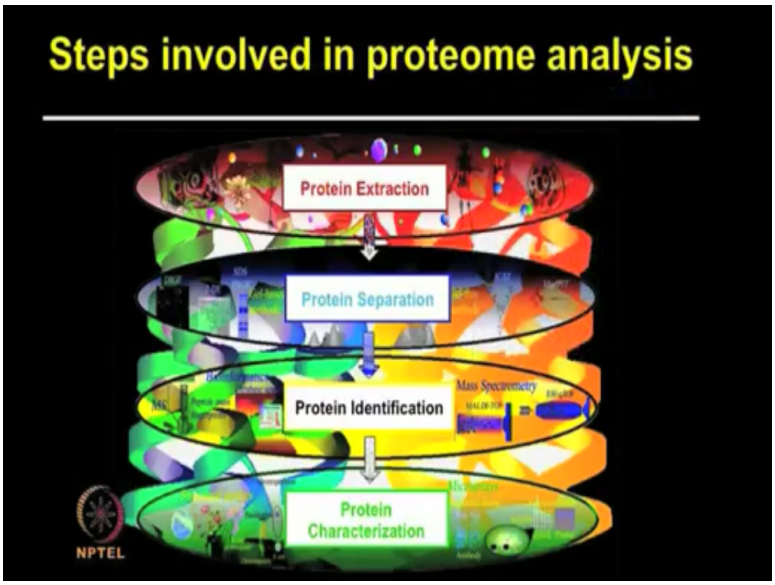
(Refer Slide Time: 00:22)



In this lecture, I will discuss about Proteomics. So first of all what is proteomics? The proteome describes the protein compliment expressed by the genome or more precisely we can say the protein complement of a given cell at a given time including the set of all the protein isoforms and its modifications. The study of entire compendium of proteins which are encoded by the genome is known as proteomics.

In this slide I have illustrated the complexity of human proteome as compared to the genome or transcriptome. The extent of diversity and complexity due to alternative splicing and post translational modifications is tremendous. Therefore, a study in proteins and proteomes are very important.

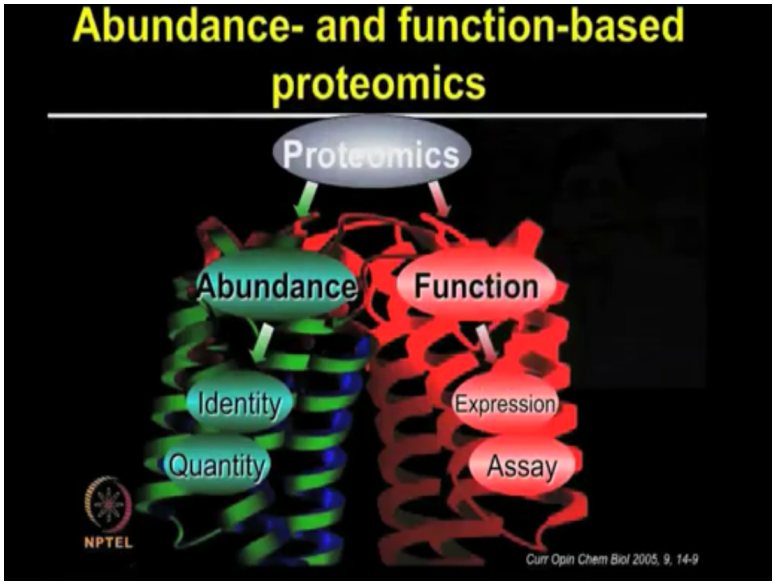
(Refer Slide Time: 01:35)



What are different steps involved in the proteome analysis? As shown here the protein extraction, protein separation, protein identification and protein characterization, these are the major steps which are involved in proteome analysis. The protein extraction from whole cells, tissue or organism is first requirement for proteome analysis.

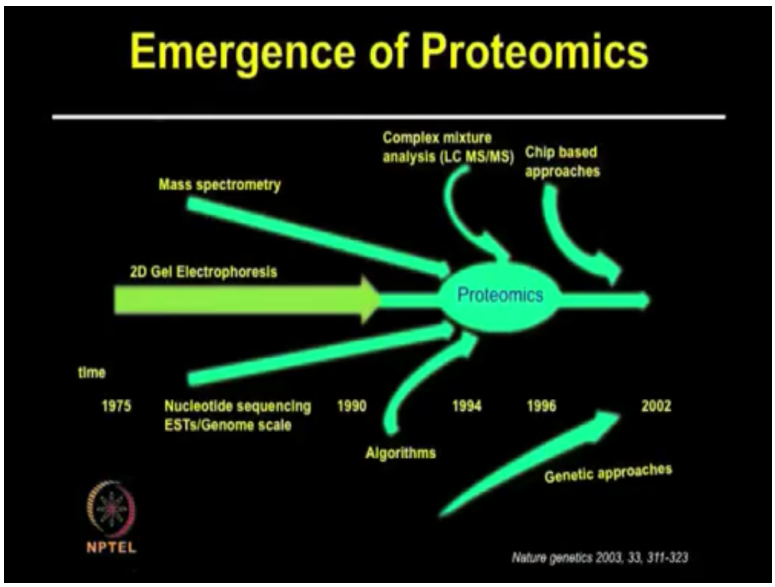
Protein separation and quantification is achieved by various proteomic techniques including gel-based techniques such as 2 dimensional electrophoresis and gel fill techniques such as (()) (02:28) mass spectrometry techniques. The functional characterization of proteins using novel proteomic platforms open new horizon for exploration in biology.

(Refer Slide Time: 02:45)



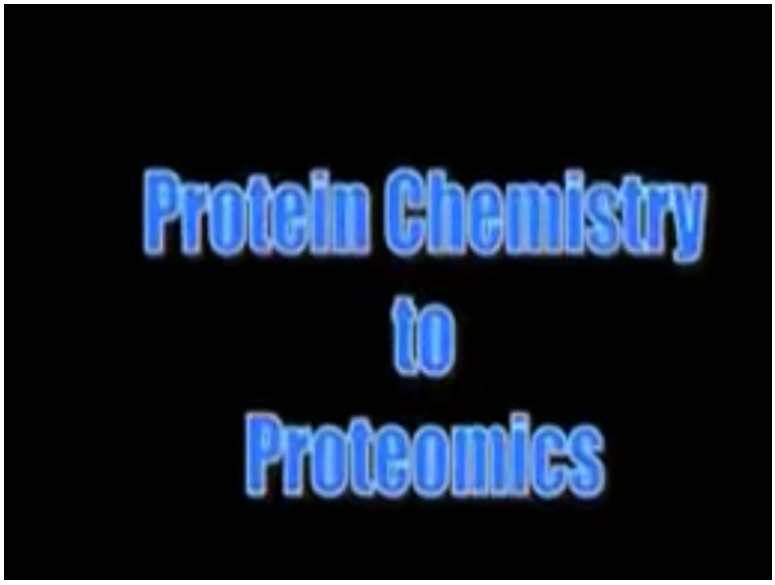
The proteomic discipline can be grouped under 2 major disciplines, abundance and function based proteomics. The abundance based proteomics aims to measure the abundance of protein extraction whereas the functional proteomics aims to determine the role of proteins by addressing protein interaction and their biochemical activities.

(Refer Slide Time: 03:16)



So how did proteomic field emerge? As you can see in the tiny slide here, shown in the slide, advancement of various techniques such as 2 dimensional electrophoresis and mass spectrometry, genome sequencing and formation and computational algorithm together led to the emergence of proteomic field.

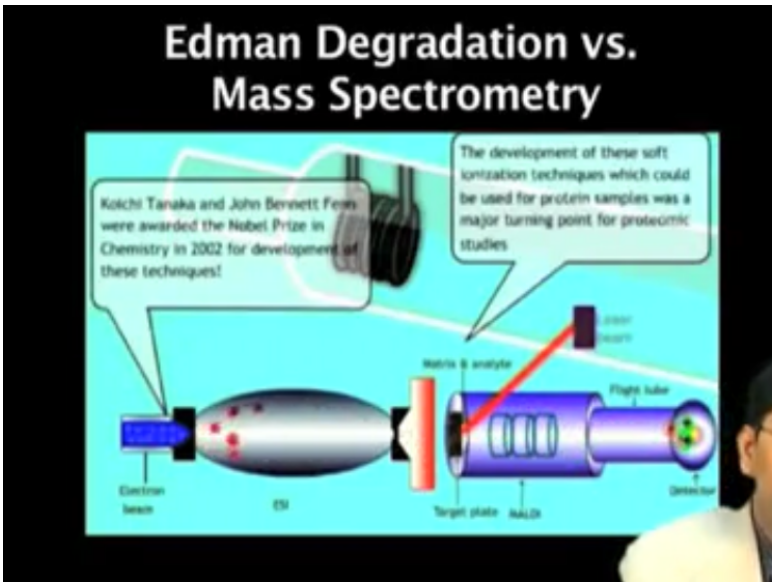
(Refer Slide Time: 03:45)



Proteomics research originates from classical protein chemistry and has embraced new heights for techniques to analyze complex samples. Many of the techniques used under the modern proteomic umbrella for example 2 dimensional electrophoresis, mass spectrometry has actually originated several years ago.

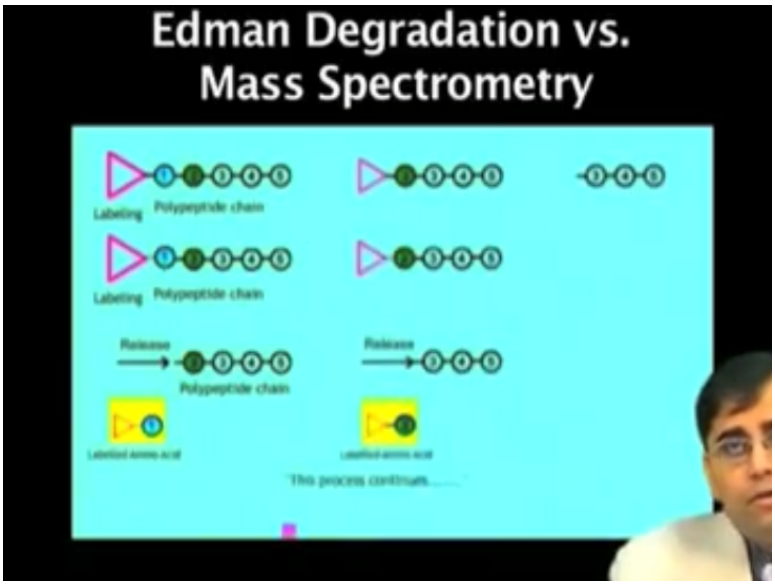
So what is new, the technological advancements in protein analysis with increased (()) (04:25), dilution and capability to carry out high throughput activities has led to the transition from protein Chemistry to new field of proteomics. Protein analysis by mass spectrometry was challenging due to complete degradation of samples with available hard ionization techniques.

(Refer Slide Time: 04:57)



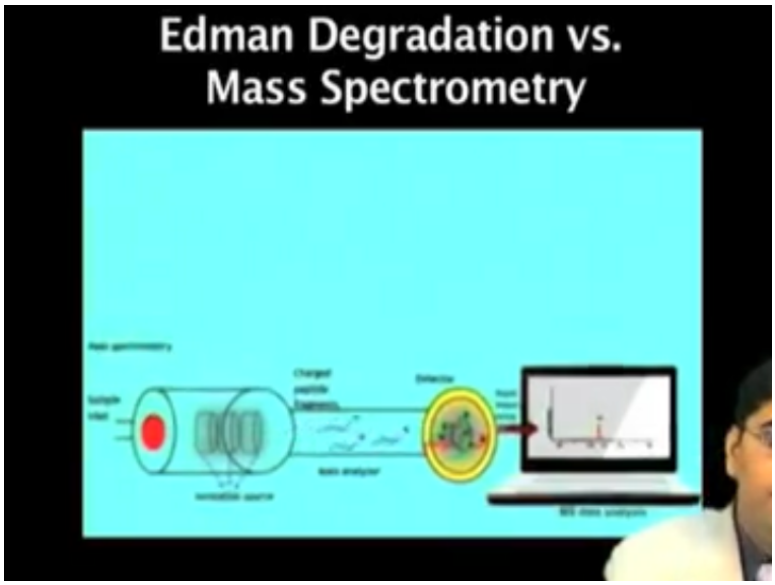
This limitation was overcome by soft ionization techniques such as (()) (05:03) and electrospray ionization. These techniques have greatly improved the proteomic studies as they facilitated mass spectrometry analysis of protein samples.

(Refer Slide Time: 05:23)



Protein sequencing by Edman degradation is time consuming and cumbersome. Several rounds of sequencing are required for analysis of polypeptide chains.

(Refer Slide Time: 05:41)

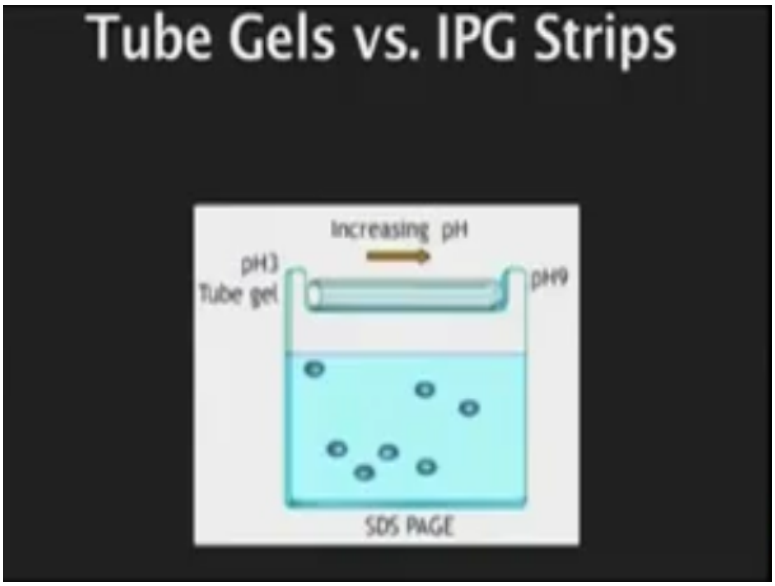


However, peptide sequencing by mass spectrometry is much faster and allows large number of samples to be analyzed in a short time.

(Refer Slide Time: 06:02)

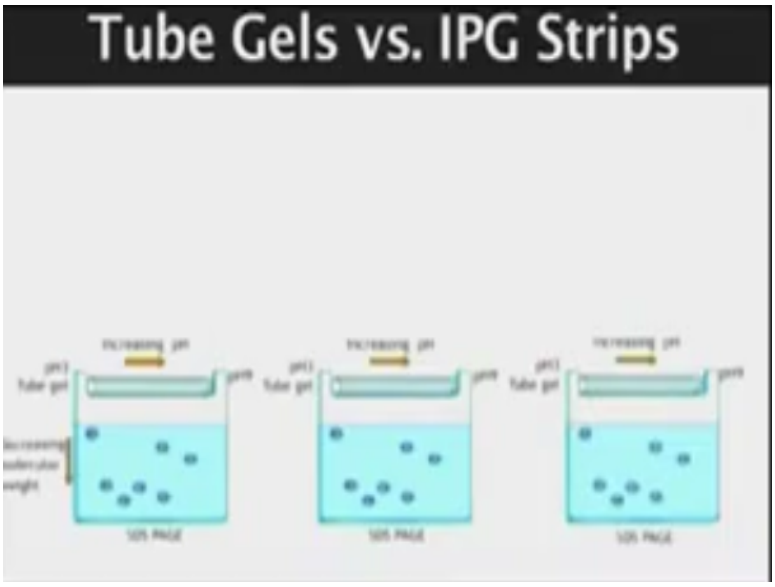
Another aspect, development of immobilized pH gradient strips facilitated proteomic analysis using 2 dimensional electrophoresis.

(Refer Slide Time: 06:24)



The pH gradient in tube gels are established by alpha light radiance which are not always very stable until to breakdown upon addition of the concentrated samples.

(Refer Slide Time: 06:28)

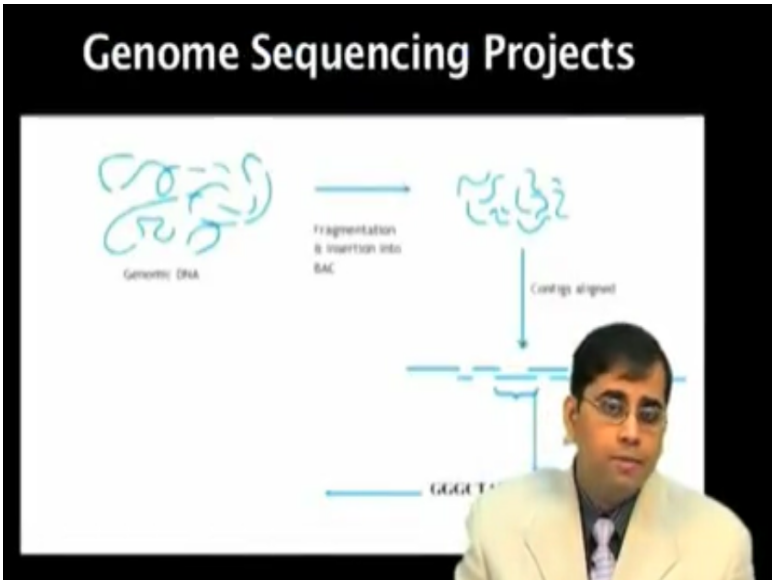


And there is a soft protein mixture by 2 dimensional electrophoresis using tube gels often resulted into variations in the gels. The problem of reproducibility was overcome to large extent by the development of immobilized pH gradient strips or IPG strips. Minimal get-to-gel variation was observed. The samples were run by 2 dimensional electrophoresis employing IPG strips which made this technique suitable for the large scale proteomic applications.

(Refer Slide Time: 07:23)

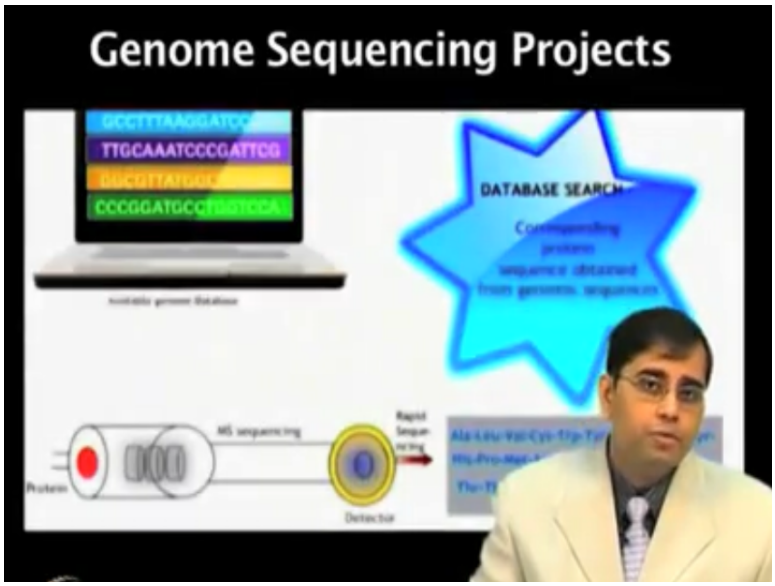
Completion of several genome sequencing projects.

(Refer Slide Time: 07:30)



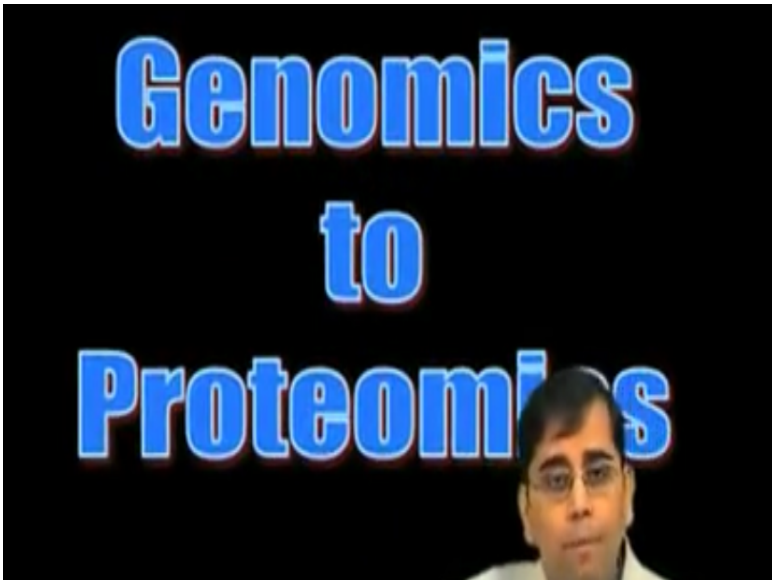
Genome sequencing of several organisms including humans have been successfully completed.

(Refer Slide Time: 07:33)



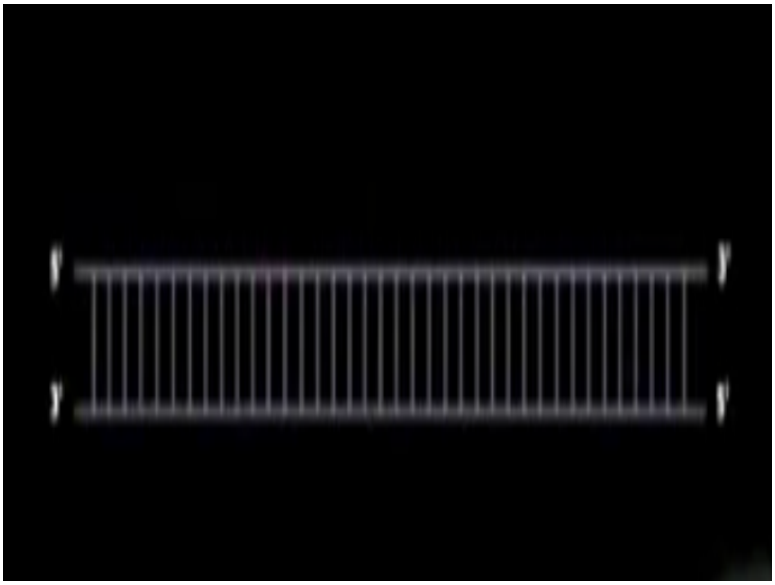
And these genome databases are extremely useful in correlation of gene and protein sequences. Several data bases are now recently available which can easily help in identification of gene sequences of your protein which has been sequenced by mass spectrometry.

(Refer Slide Time: 08:10)



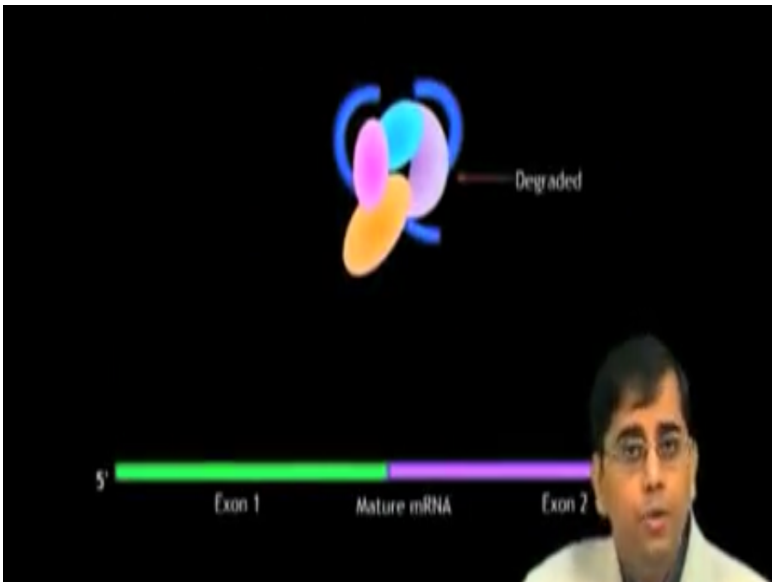
Genome represents an important starting point towards understanding complexity of biological functions, however, proteins provide a much more meaningful insight into the mysteries of essential biological processes. To obtain better understanding of similar processes and regulation there has been an increasing interest in studying proteome. There are several reason why one need to study proteomics.

(Refer Slide Time: 08:54)



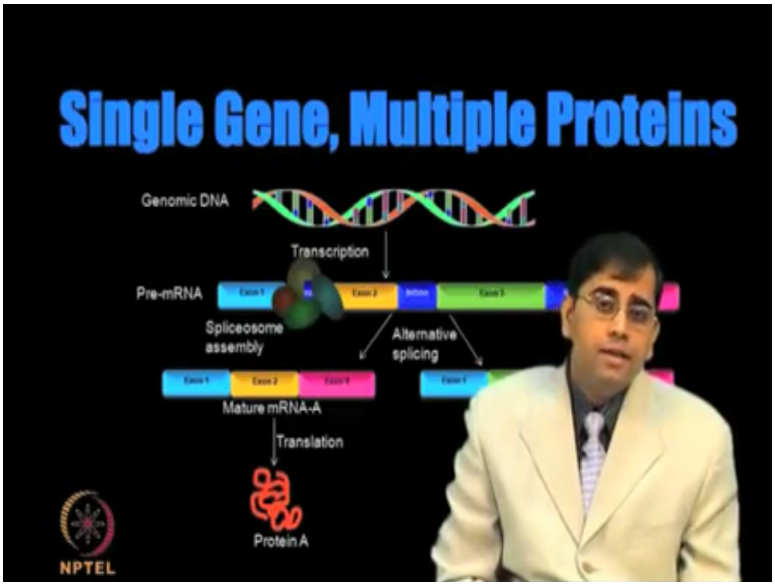
First, the genomic DNA contains large strictures of noncoding regions.

(Refer Slide Time: 09:00)



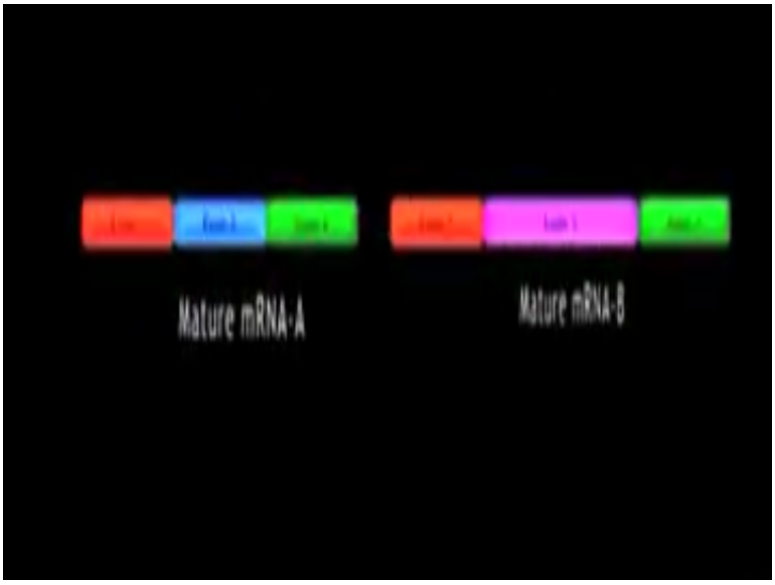
The pre mRNA is synthesized from the genomic DNA by the process of transcription. mRNA contains both exons, the coding sequences as well as introns which are intervening non coding sequences. By involving series of steps, finally the free 3 prime hydroxyl group of the first exon attacks the 5 prime end of the second exon such that they are joined together to give the matured mRNA.

(Refer Slide Time: 09:51)



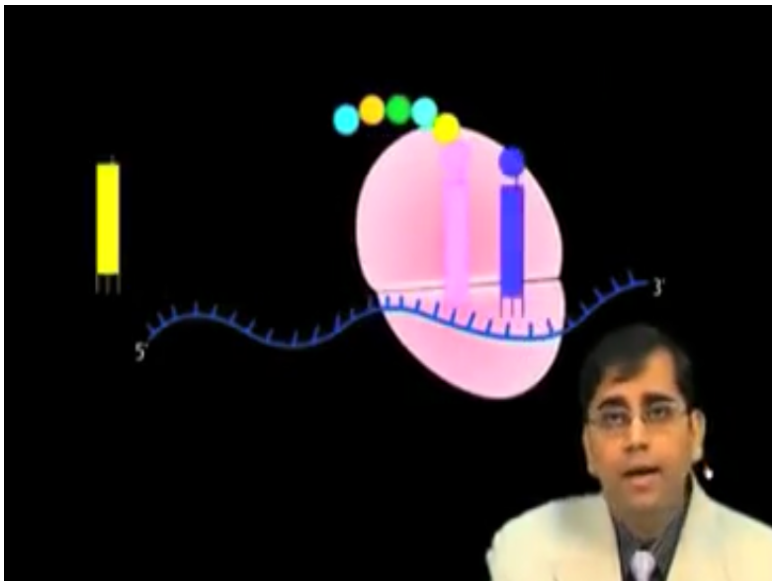
Second important factor is - single gene can give rise to multiple proteins.

(Refer Slide Time: 09:57)



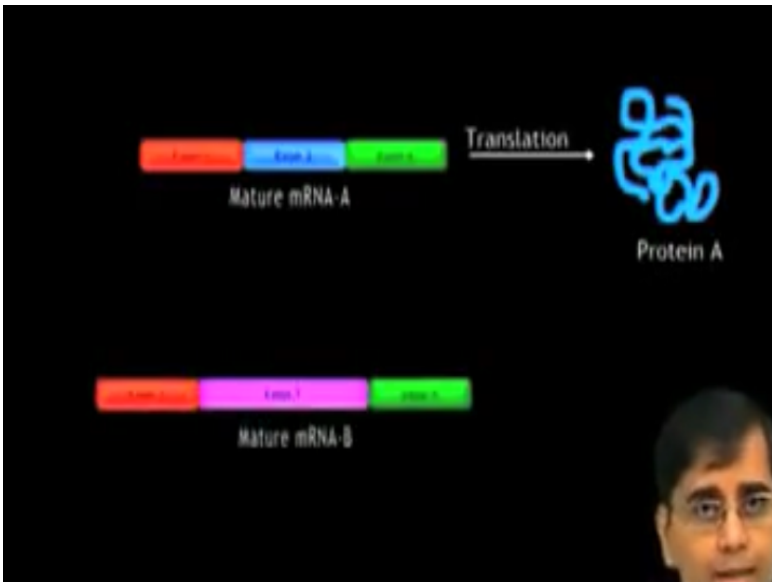
The alternative splicing is a process by which exons or coding sequences of pre mRNA produced by transcription of a gene are combined in different ways during omni splicing.

(Refer Slide Time: 10:22)



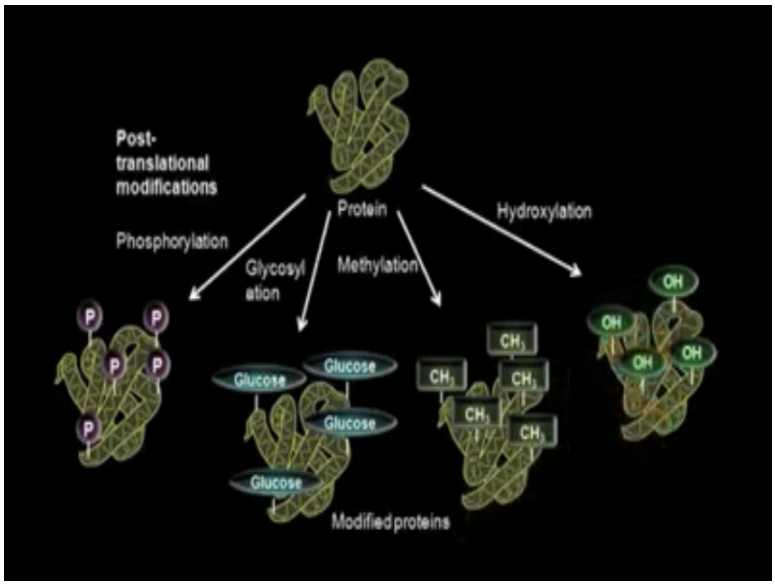
Resulting matured mRNA give rise to different protein products by translation, most of which are isoforms of one another.

(Refer Slide Time: 10:37)



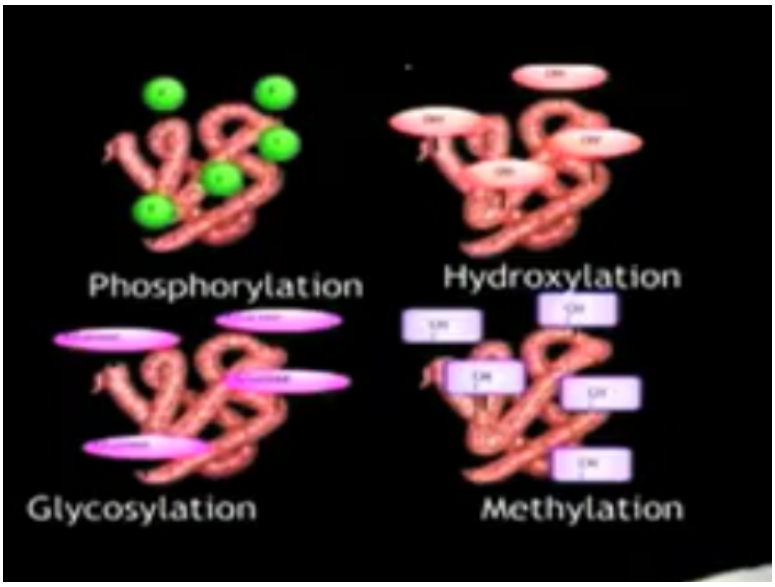
The diversity of proteins encoded by a genome is greatly increased due to alternative splicing.

(Refer Slide Time: 10:47)



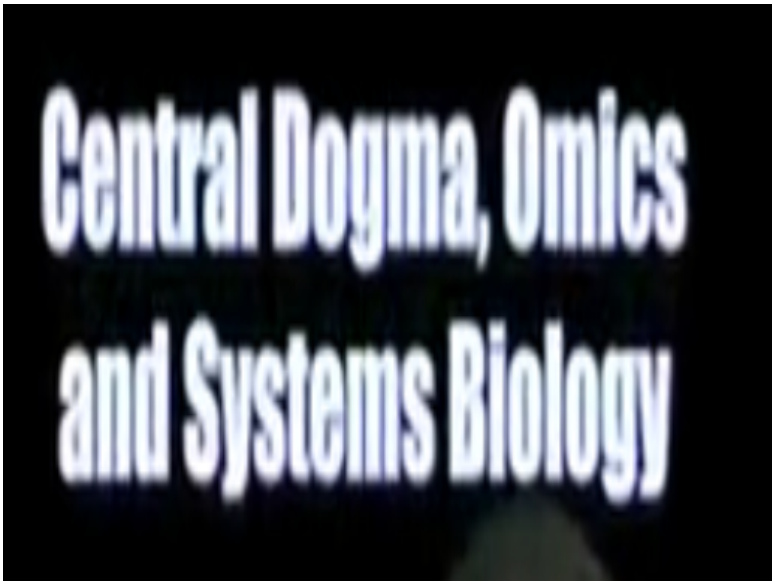
Third important factor is post translational modification of proteins.

(Refer Slide Time: 10:56)



The proteins obtained by translation undergoes folding and various post translational modifications such as phosphorylation, glycoxylation, alkylation, hydroxylation etc. to give the final functional protein. The PTMs generate diversity, complexity and heterogeneity of gene products and its functional consequences can be modulation in protein dynamics and alteration of its functional activity.

(Refer Slide Time: 11:45)

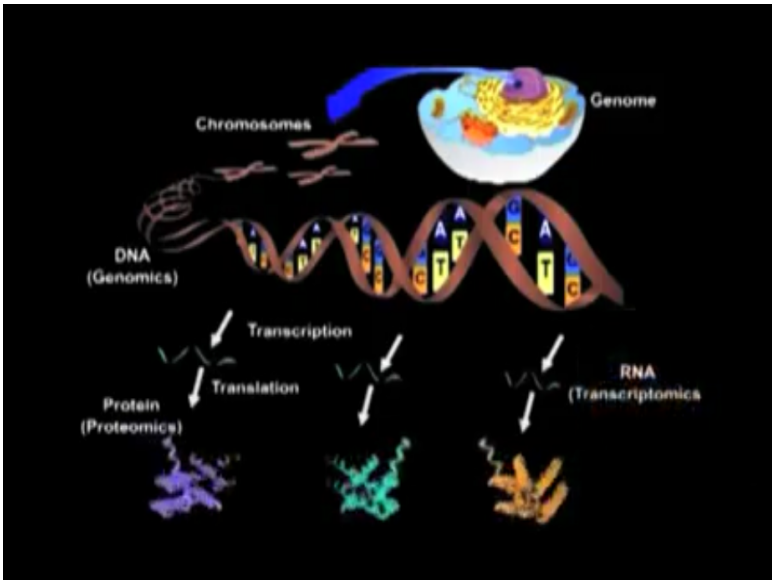


During the last decade we have witnessed the regulation in biology as this discipline has fully embraced omics to its emergence of genome wide analysis to understand cellular DNA, RNA and protein content by employing genomics, transcriptomics and proteomics at systems level has revolutionized our understanding of control networks that mediate this cellular processes. These concepts will be discussed in first module.

(Refer Slide Time: 12:34)

Genes are the blueprint for life and proteins are the effector molecules.

(Refer Slide Time: 12:40)

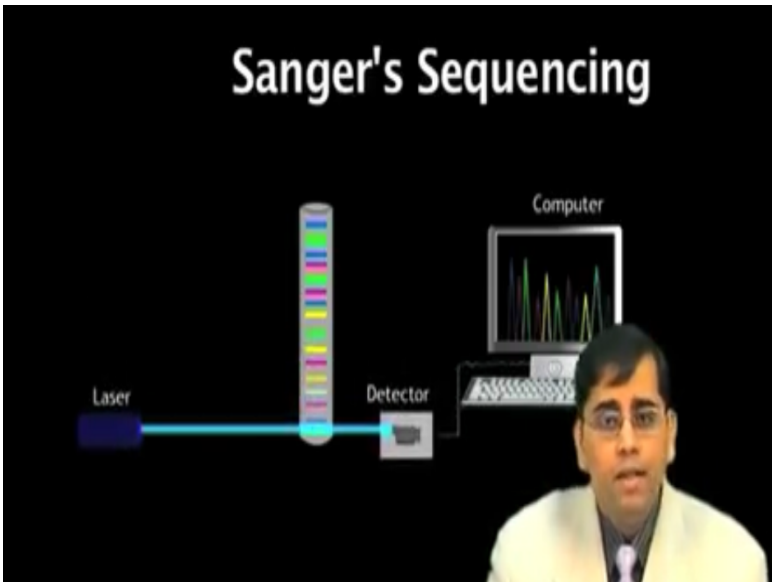


Due to this fact, the central dogma has guided research at the systems level. After completion of human genome sequence, number of genes around 25,000 are surpassed by an estimated number of proteins in millions, studying large scale study of protein structure and function requires a thorough understanding of protein composition and their various structural levels by employing high throughput tools.

(Refer Slide Time: 13:30)

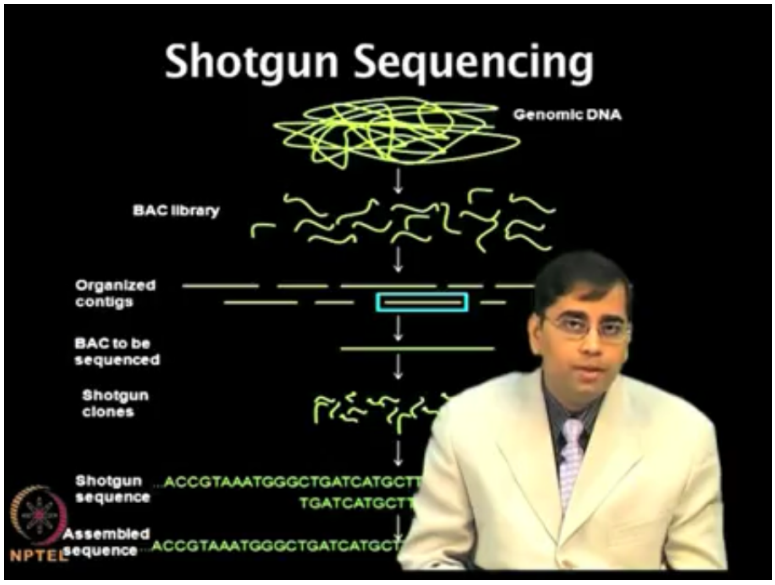
You study genome of an organism by employing sequencing and genome mapping is known as genomics.

(Refer Slide Time: 13:41)



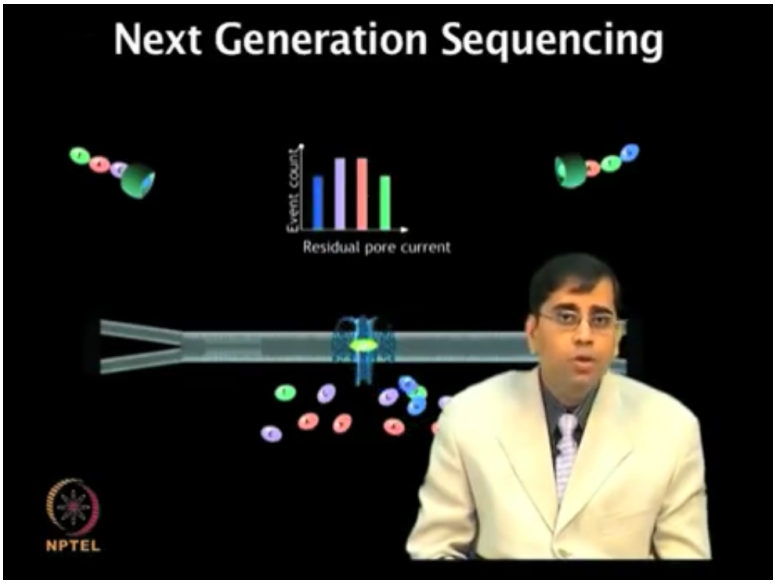
Several genome sequencing projects that aim to elucidate the complete genome sequence of organisms have been undertaken by several research groups all over the world.

(Refer Slide Time: 13:56)



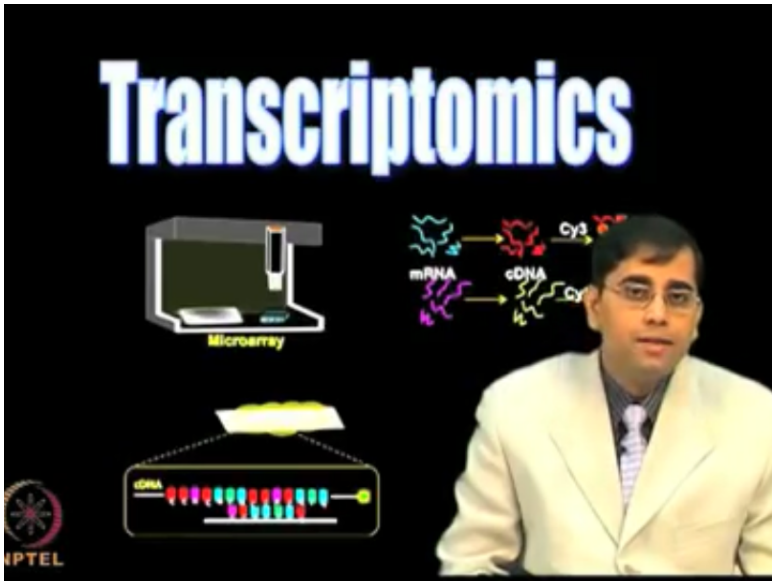
From a genomic library clones were isolated and ordered into a detailed physical map. Further individual clones were sequenced by shotgun sequencing to provide the complete genome sequence.

(Refer Slide Time: 14:41)



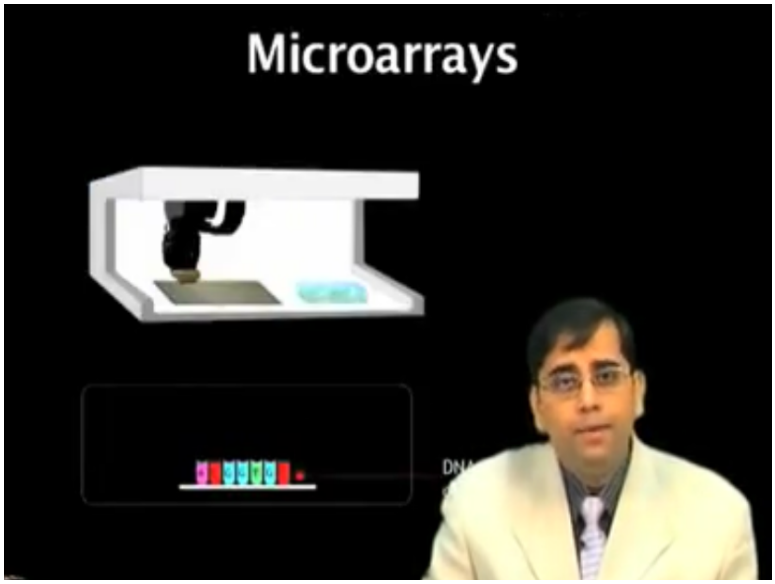
Recently net generation sequencing NGS (()) (14:21) have dramatically increased the pace of sequencing by several order of magnitudes. Next generation sequencing based on Nano-pore structures is known as Nano-pore sequencing.

(Refer Slide Time: 14:45)



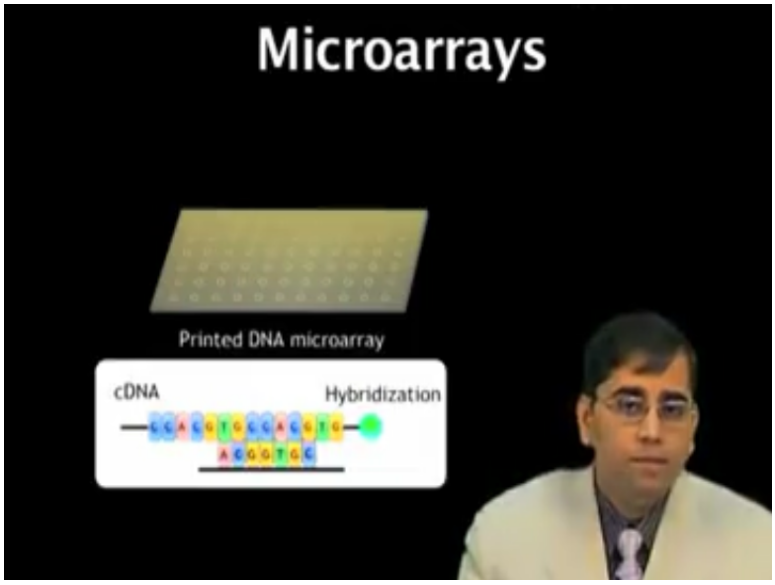
Transcriptomics.is study of all the mRNA molecules expressed by a particular cell type of an organism is known as transcriptomics. The transcriptomics analysis measures the genes that are being actively expressed at any given time and varies significantly with external environmental conditions. Various techniques such as microarrays, QRT PCR etc. have been widely used for transcriptional analysis.

(Refer Slide Time: 15:27)



In a micro array experiment, the mRNA from control and test samples are extracted and reverse transcribed into its corresponding cDNA.

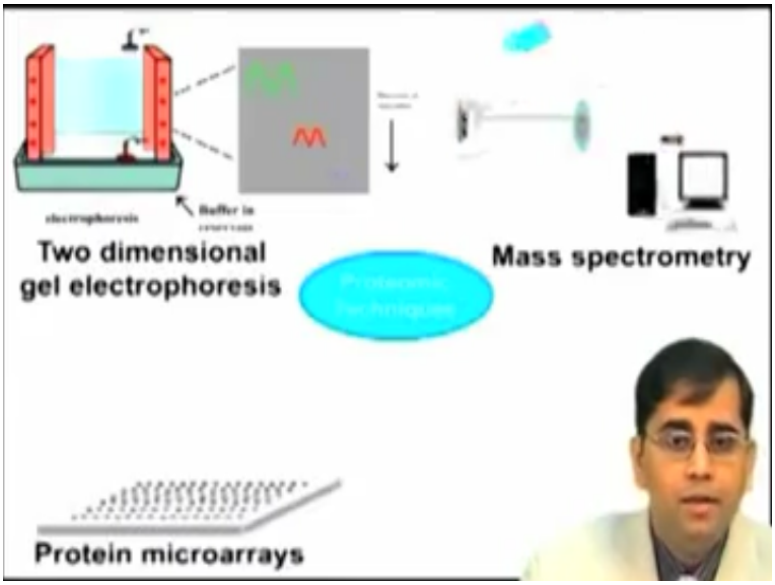
(Refer Slide Time: 15:39)



The cDNA samples are labeled with (()) (15:46) dice and mixed cDNA sample is incubated on printed DNA microarray. This allows hybridization to occur between the pro oligoneutriotide on array surface and the labeled cDNA sample of interest. In this manner expression level of 1000s of genes can be measured and analyzed simultaneously.

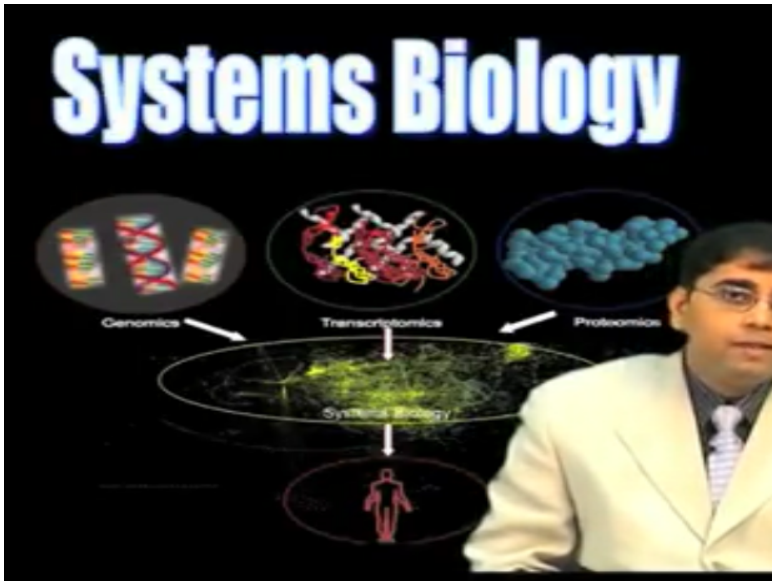
(Refer Slide Time: 16:25)

(Refer Slide Time: 16:28)



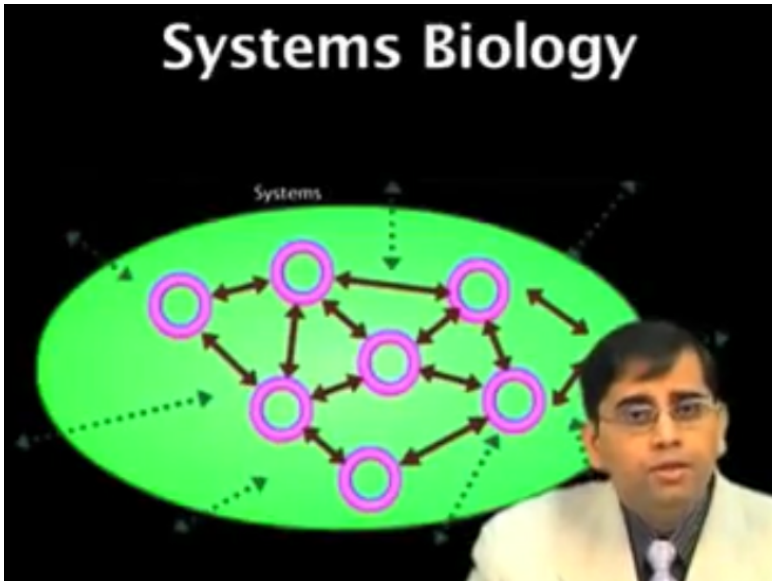
Different types of proteomic technologies such as 2 dimensional electrophoresis, mass spectrometry, microarrays and label fill techniques which will be discussed in more details later.

(Refer Slide Time: 16:44)

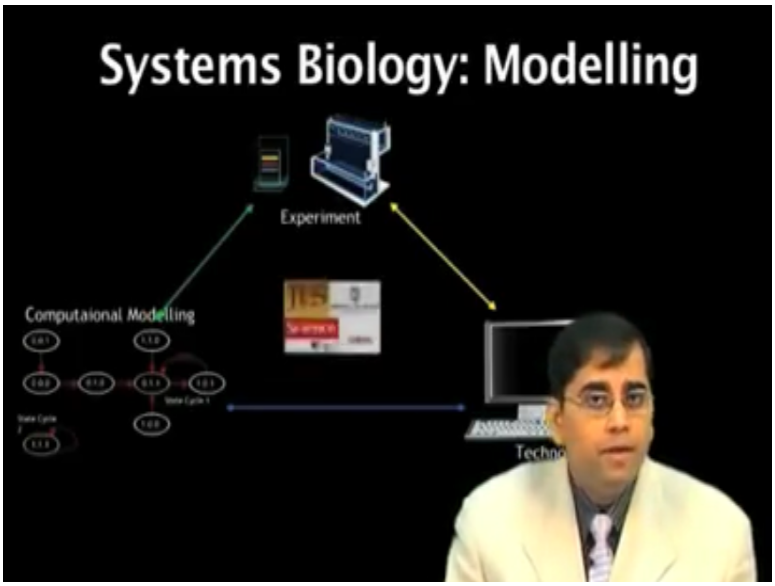


In omics era, technological advancement in genomics, proteomics and metabulomics have generated large scale datasets in all the aspects of biology. These large datasets have motivated the computational biology and systems approaches with objective of understanding the biological system as a whole.

(Refer Slide Time: 17:16)

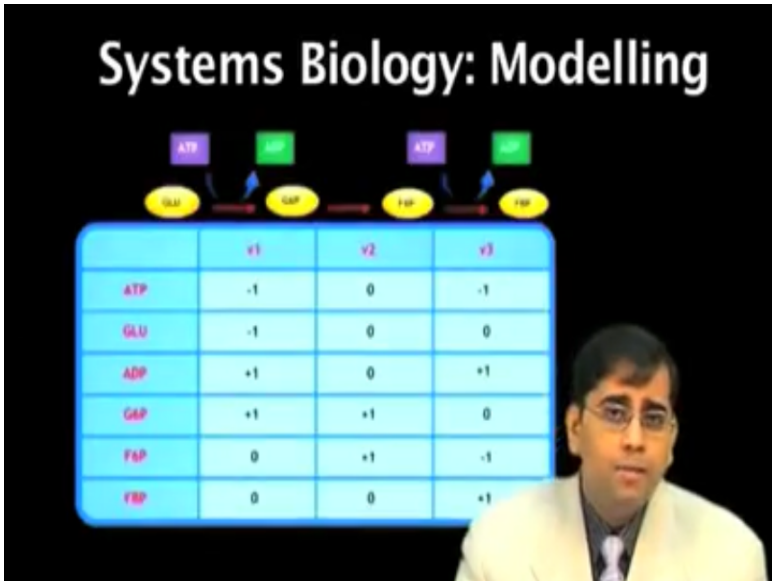


(Refer Slide Time: 17:31)

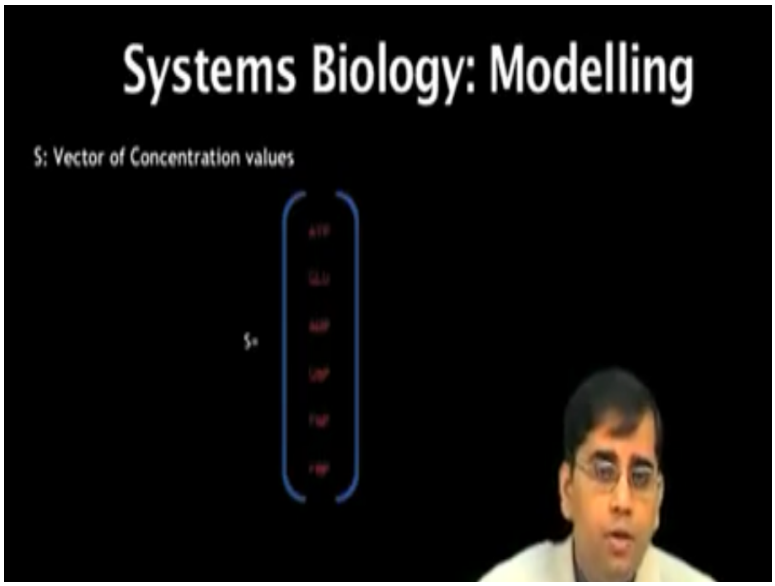


The system biology and biological network modeling aims to understand the biological processes as whole system rather than the isolated parts by synergistic application of experiment theory, technology and modeling.

(Refer Slide Time: 17:38)

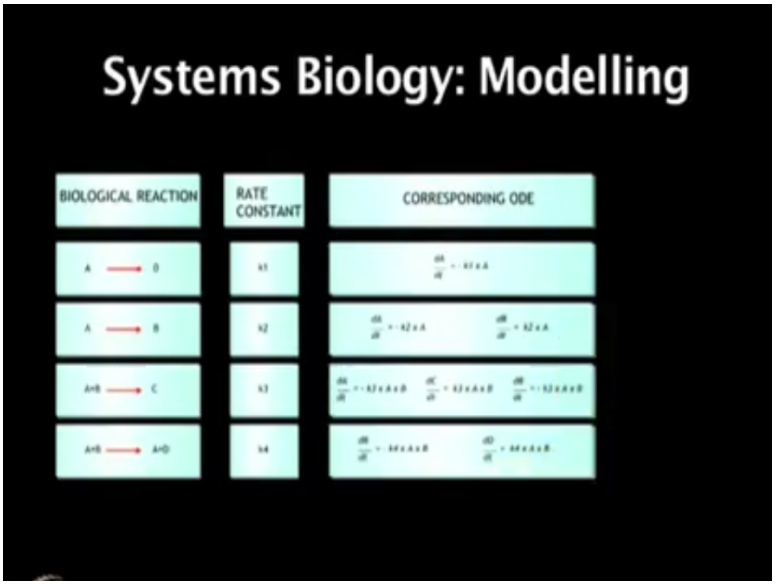


(Refer Slide Time: 17:49)



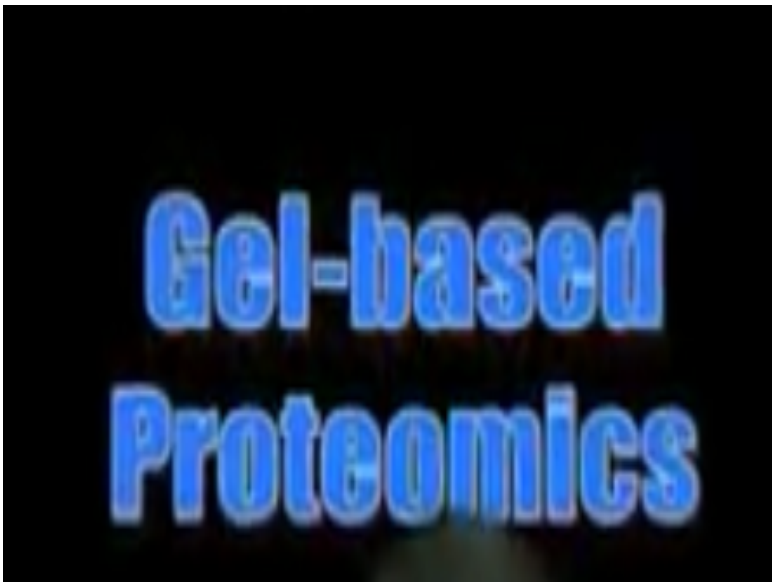
The system level studies aim to develop competitionally efficient and reliable models of underlying gene regulatory networks.

(Refer Slide Time: 17:51)



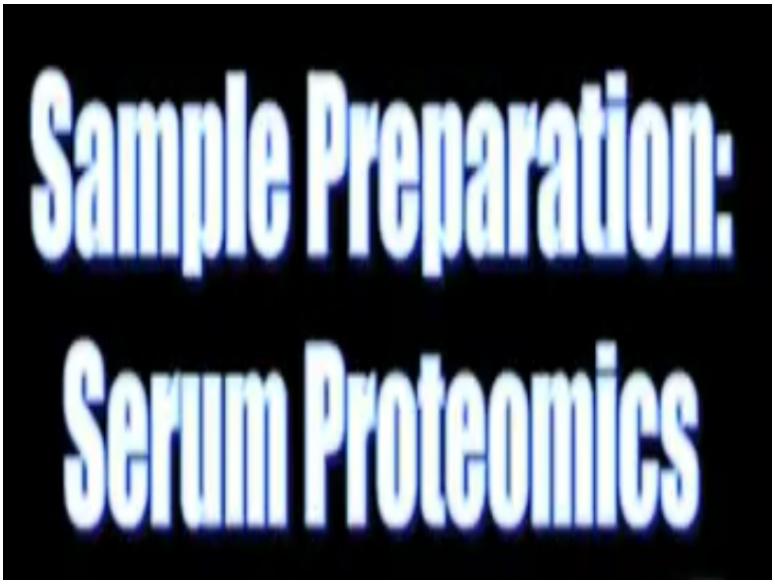
The quantitative analysis measures and aims to make models for precise (()) (17:59) parameters of a systems network component. It also uses properties of network connectivity.

(Refer Slide Time: 18:16)

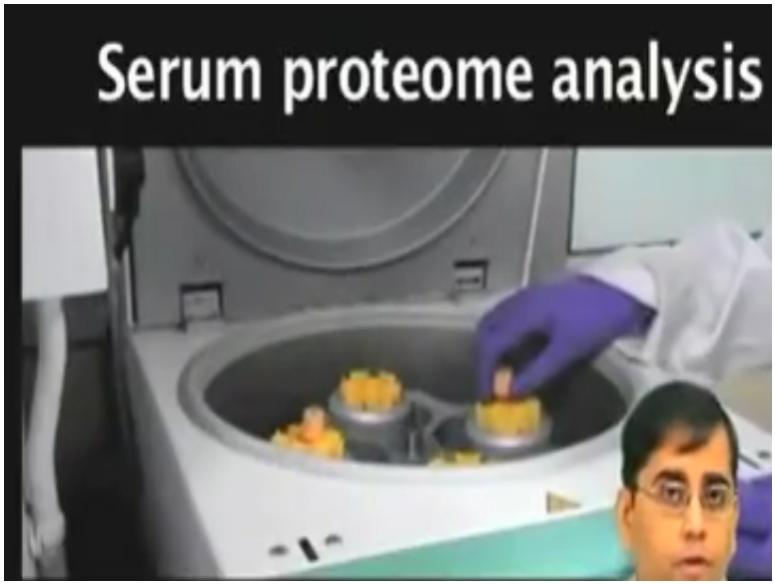


Several techniques used in proteomics typically aim to elucidate the expression, localization, interaction and cellular function of proteins. SDS page 2 dimensional electrophoresis, difference in the electrophoresis are various commonly used gel based proteomic techniques. Protein extraction is the first step for the proteomic analysis. The protein extraction method aims that most, if not all the proteins in a cell or its organelles are extracted by the procedure in the presence of interfering components are reduced or minimized.

(Refer Slide Time: 19:14)

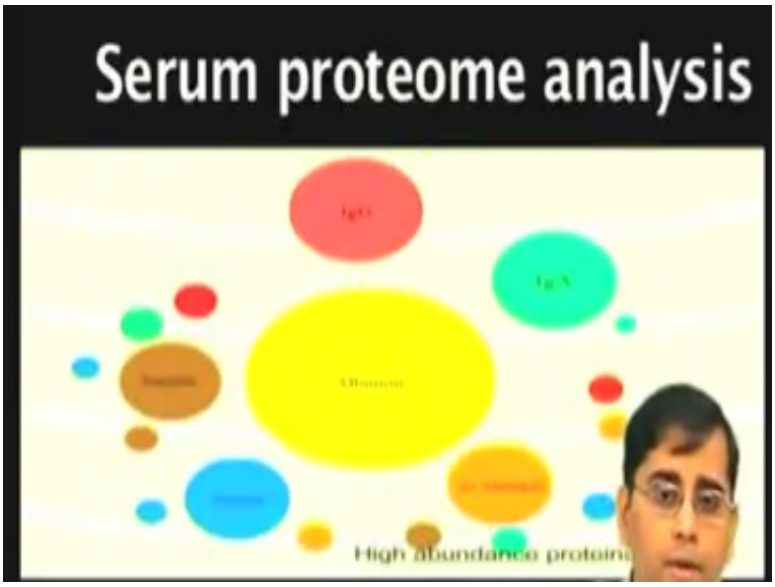


(Refer Slide Time: 19:39)

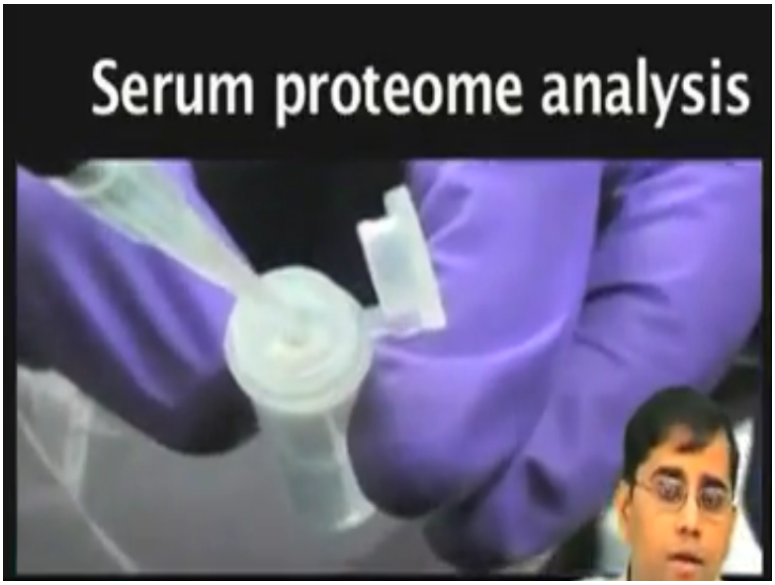


Different biological systems, different biological samples pose different type of challenges. For example, serum protein analysis shown here illustrates that proteins in biological system such as serum may have difference of several order of magnitude.

(Refer Slide Time: 19:47)

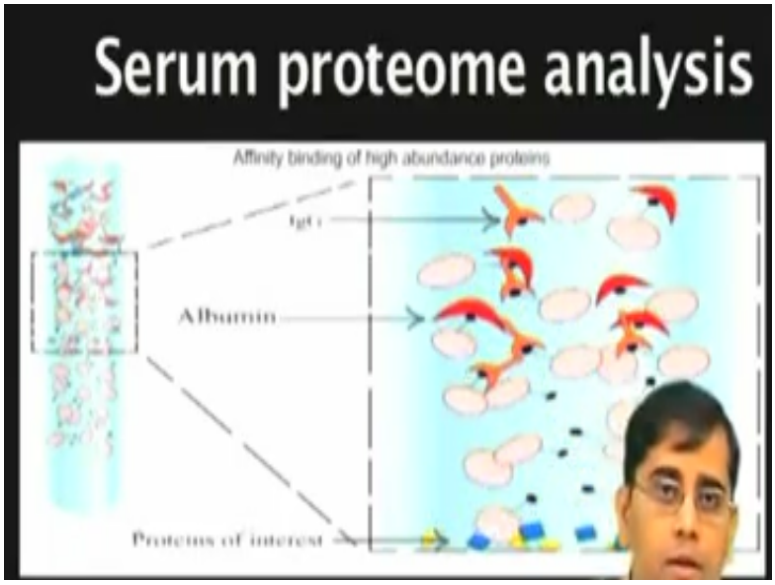


(Refer Slide Time: 19:52)



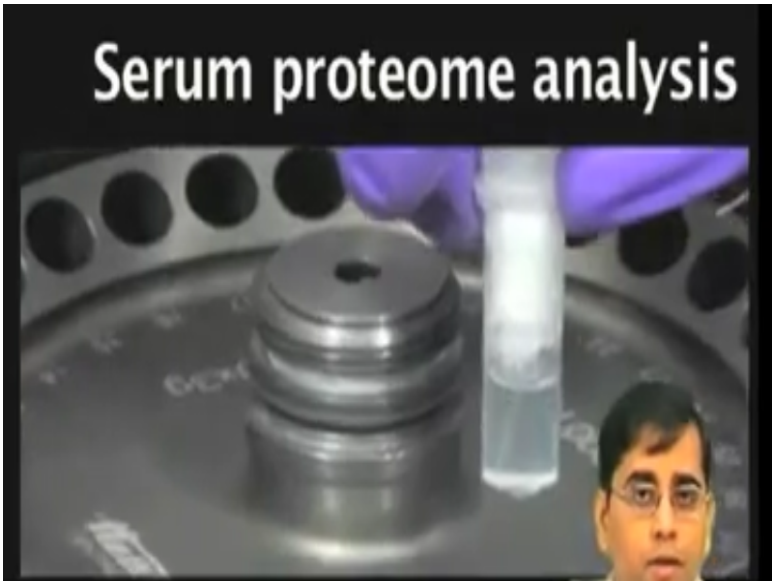
Albumin and immunoglobulin are the most abundant proteins in serum which mask other low abundant proteins which are present in the lower concentration.

(Refer Slide Time: 19:58)

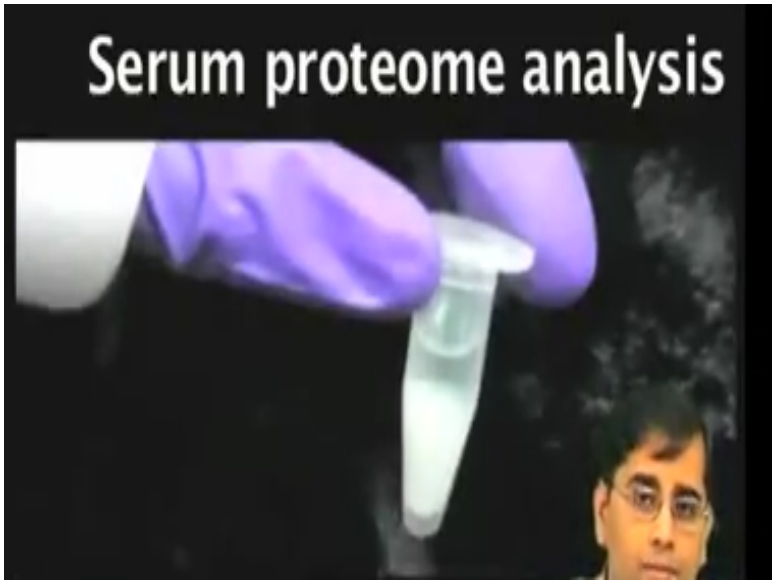


It is therefore preferred to remove these high abundance proteins by using affinity chromatographic methods.

(Refer Slide Time: 20:13)

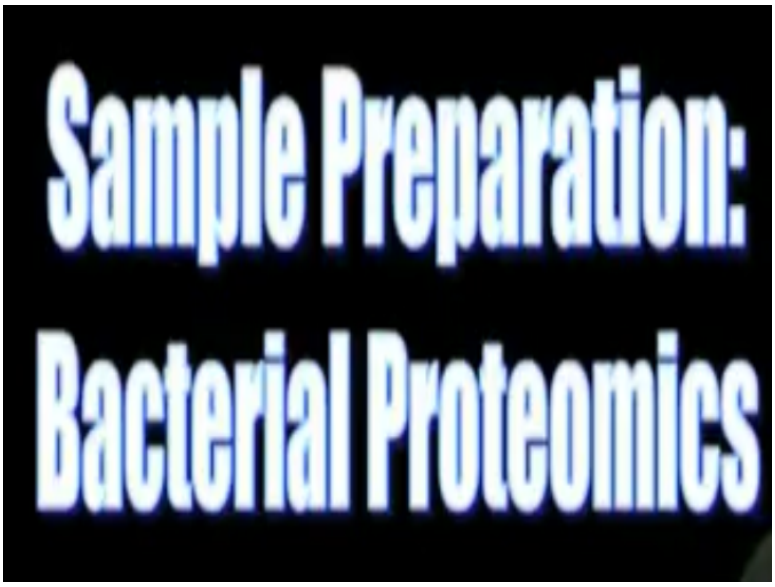


(Refer Slide Time: 20:14)



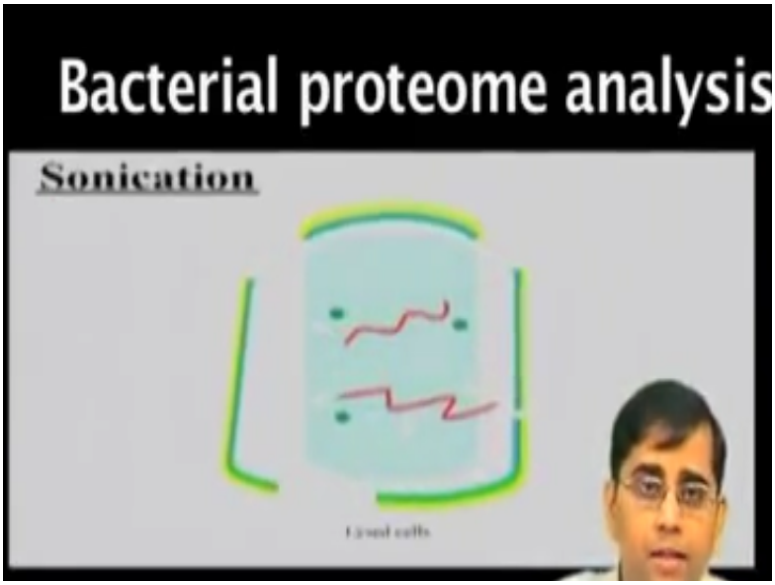
While the serum has been processed using this chromatographic methods, these proteins can be extracted.

(Refer Slide Time: 20:20)



In bacterial protein sample preparation sonication is an important step to disrupt the bacterial membrane

(Refer Slide Time: 20:32)



Sonication breaks open the cellular content, the cellular membranes (()) (20:36) the intracellular contents

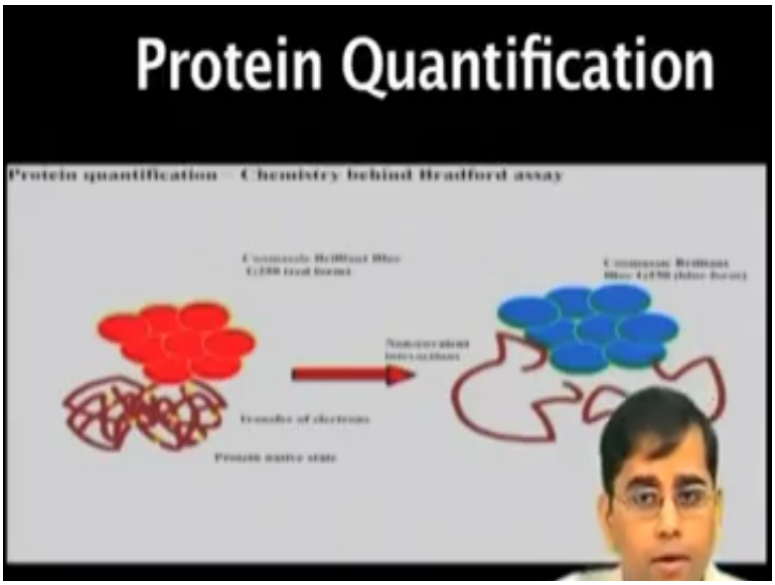
(Refer Slide Time: 20:41)

Protein extraction can be performed by using different methods and protein pellets are reconstituted in lysis buffer for proteomic analysis.

(Refer Slide Time: 20:59)

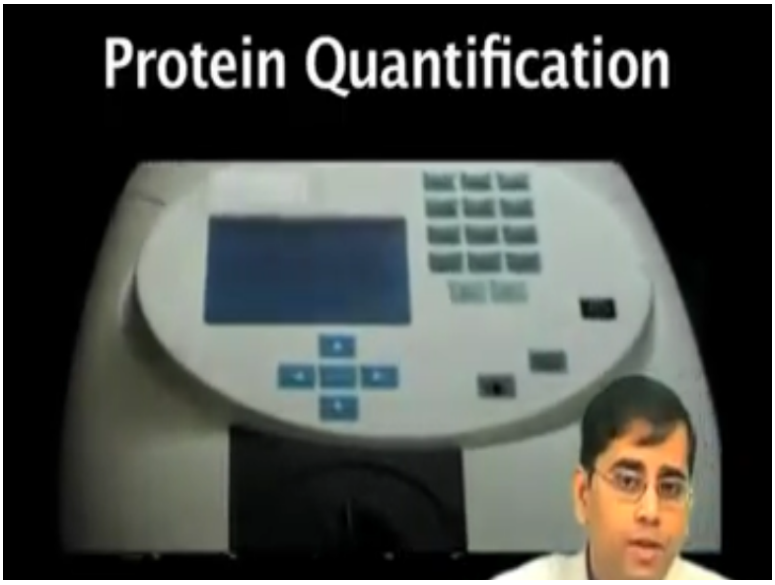
Protein quantification is sensitive to detergents or certain ions. Therefore, it is crucial to sell it the correct quantification method.

(Refer Slide Time: 21:15)

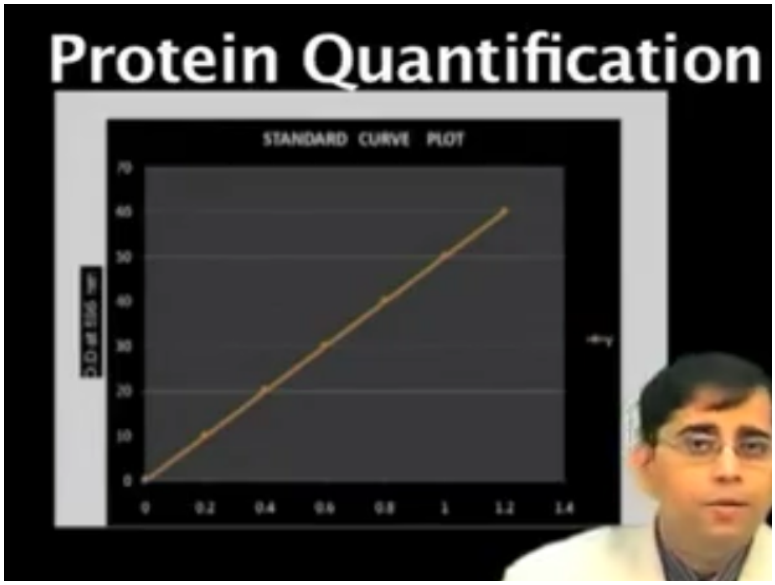


In Bradford color agent, transfer of electron converts the dye to its blue form thereby giving the solution blue color.

(Refer Slide Time: 21:29)



(Refer Slide Time: 21:43)

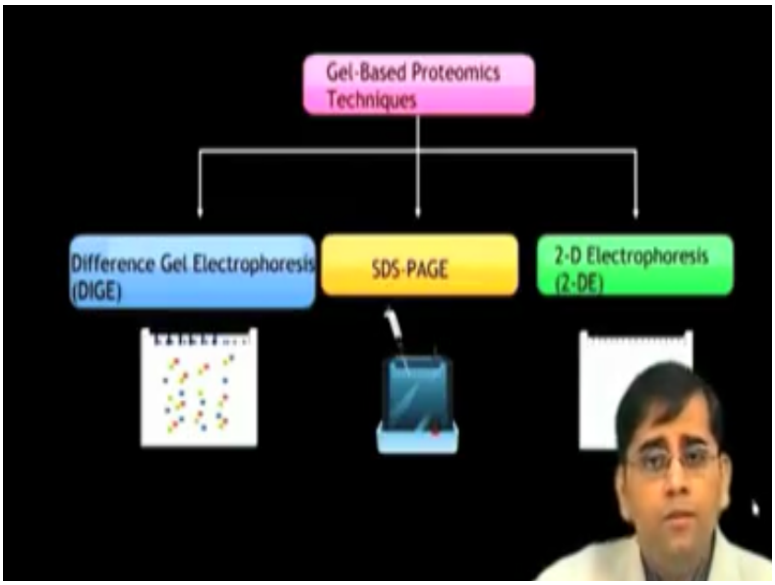


Absorbance of standard and unknown protein samples can be measured at 595 nanometers and protein concentration can be determined from the standard plot of the absorbance values.

(Refer Slide Time: 21:46)

In gel based proteomics, proteins are commonly analyzed using SDS page and 2 dimensional gel electrophoresis.

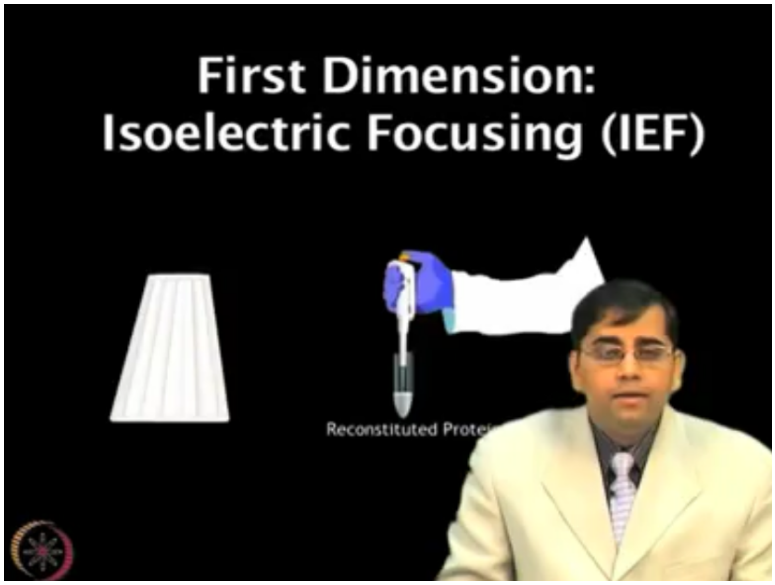
(Refer Slide Time: 21:57)



Separation in SDS page occurs almost excessively on the basis of molecular weight whereas in 2-DE the first dimensional separation is based on isoelectric point and second dimensional separation is based on molecular weight. Some of the limitations of 2 dimensional electrophoresis can be overcome by difference, gel electrophoresis or DIGE technique. 2-DE or DIGE in comparison with mass spectrometry has been the standard technique for proteome analysis.

The 2 dimensional electrophoresis involves protein separation on a pH gradient based on their isoelectric point using isoelectric focusing followed by separation in second dimension using SDS page.

(Refer Slide Time: 23:12)



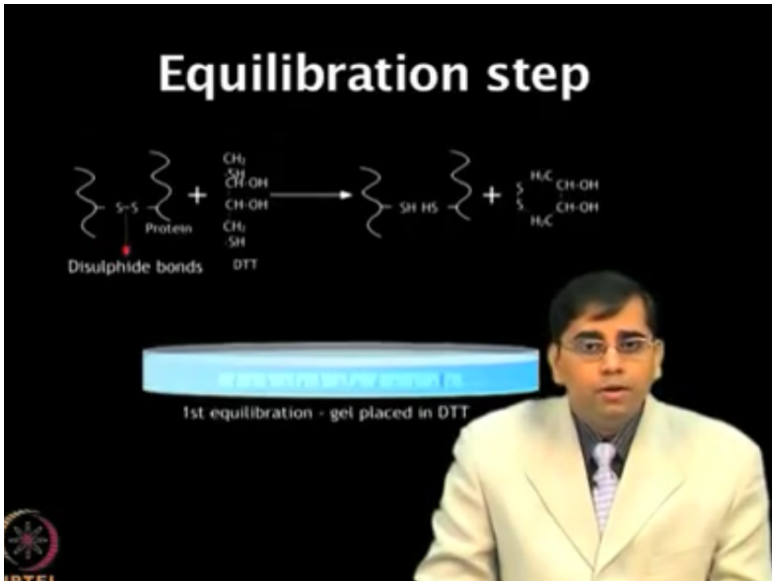
To perform 2-DE at the reconstituted protein sample to the rehydration trend and place the IPG strip for rehydration.

(Refer Slide Time: 23:24)

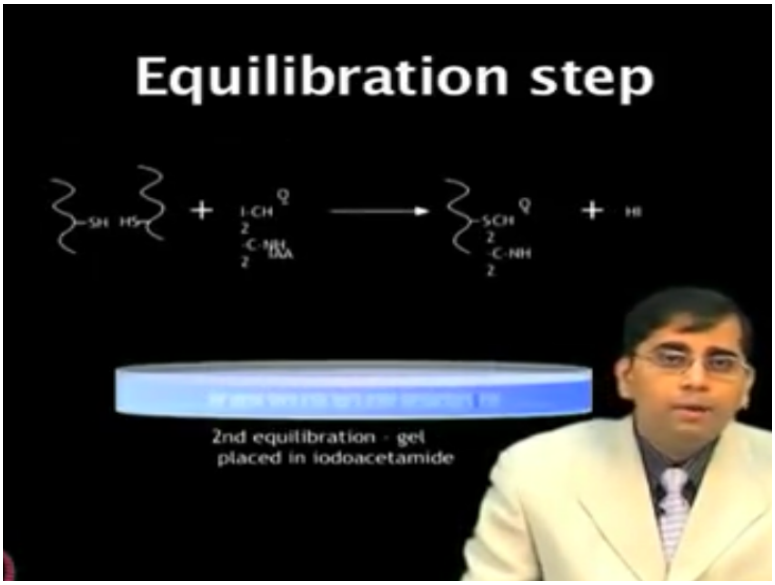
(Refer Slide Time: 23:43)

Isoelectric focusing involves the application of an electric field which causes the proteins to migrate to the position on the pH gradient strips that matches the PI of a specific protein after which it does not move in the electric field owing to the lack of charge, the proteins migrate along the strip and come to a rest at a point where the net charge becomes 0 known as isoelectric point.

(Refer Slide Time: 24:08)

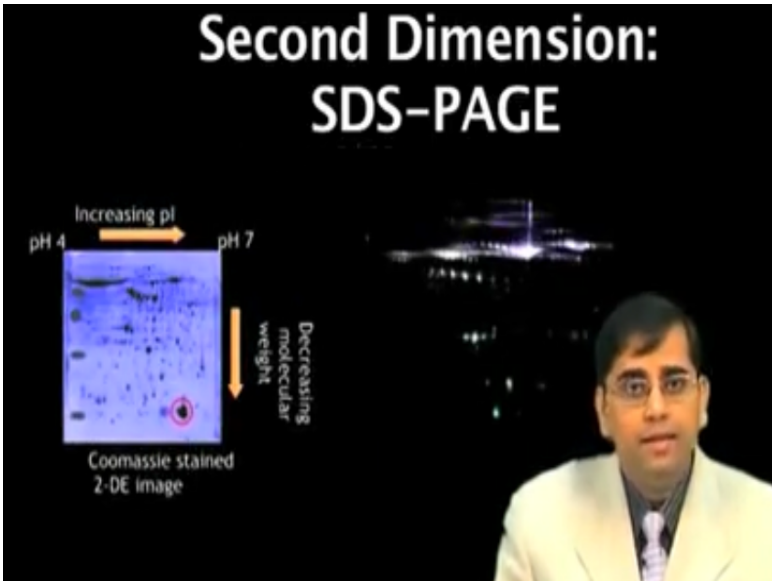


(Refer Slide Time: 24:20)



Prior to the second dimensional separation an equilibration step is required. In equilibration Dithiothreitol brings about cleavage of the protein disulphide bonds while iodoacetamide prevents reformation of these bonds by buying into the self hydral groups.

(Refer Slide Time: 24:35)



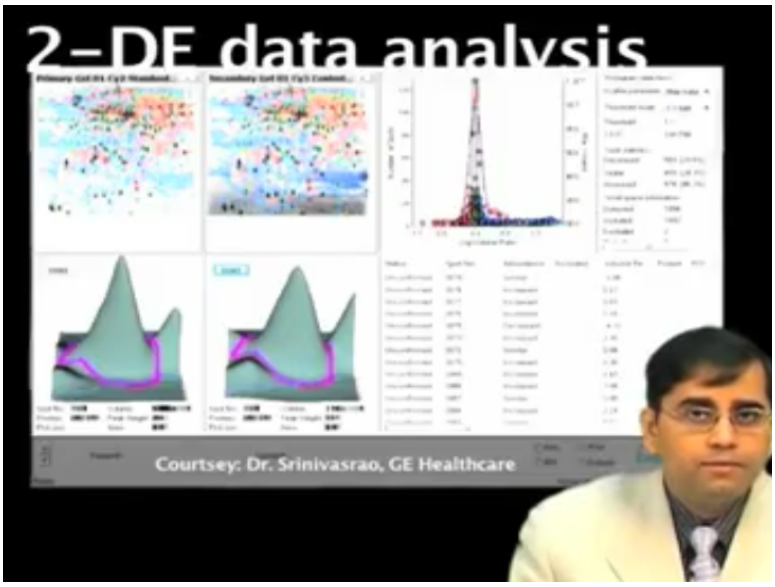
On SDS page gel, proteins get separated on the basis of their molecular weight with a low molecular weight proteins have high mobility and migrating further through the gel and the high molecular weight protein remains close to the point of sample application. Gels can be visualized by different staining methods such as Coomassie staining, silver staining and staining DIGE.

(Refer Slide Time: 25:13)

(Refer Slide Time: 25:16)

The gel data analysis will be discussed with an application expert of GE Healthcare.

(Refer Slide Time: 25:21)

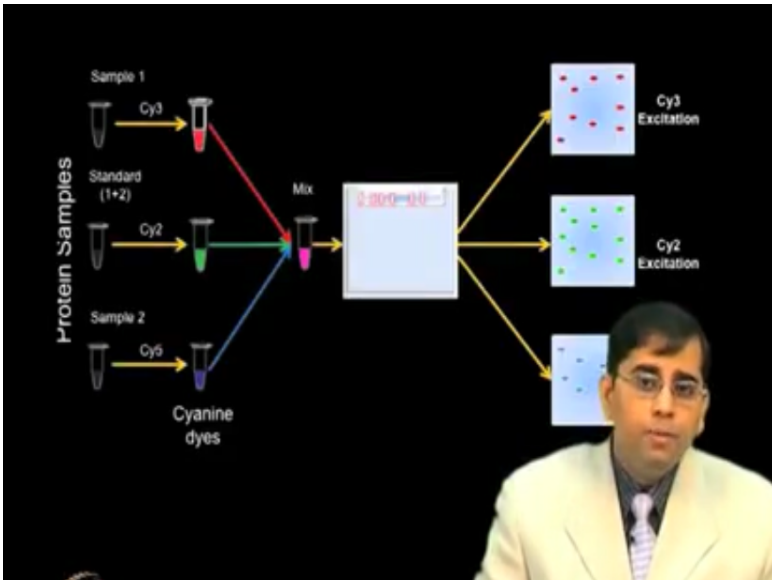


The gel analysis involves image processing, detection of spots, making (()) (25:30), land marking viewing histograms etc. Various information regarding the spots such as the area, volume intensity and statistical parameters such as standard deviations can also be calculated.

(Refer Slide Time: 25:57)

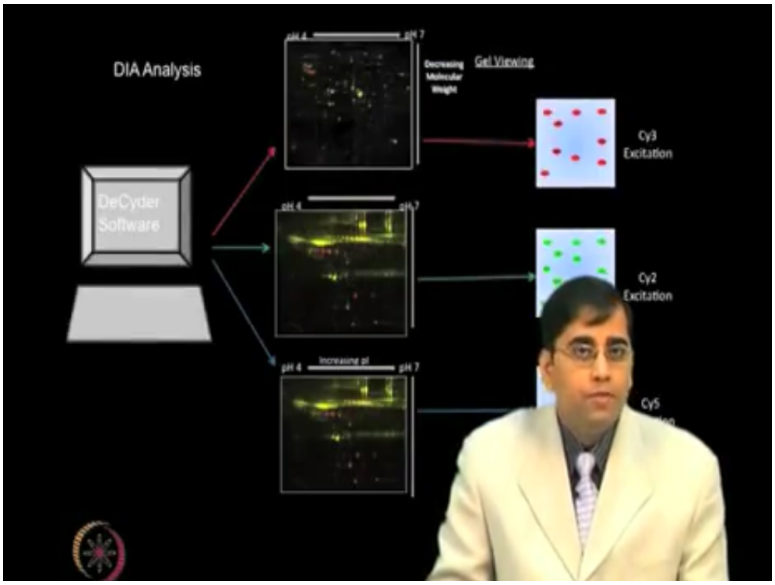
2 dimensional electrophoresis has high resolving power but it has several limitations such as staining artifacts and reproducibility in gel to gel.

(Refer Slide Time: 26:10)



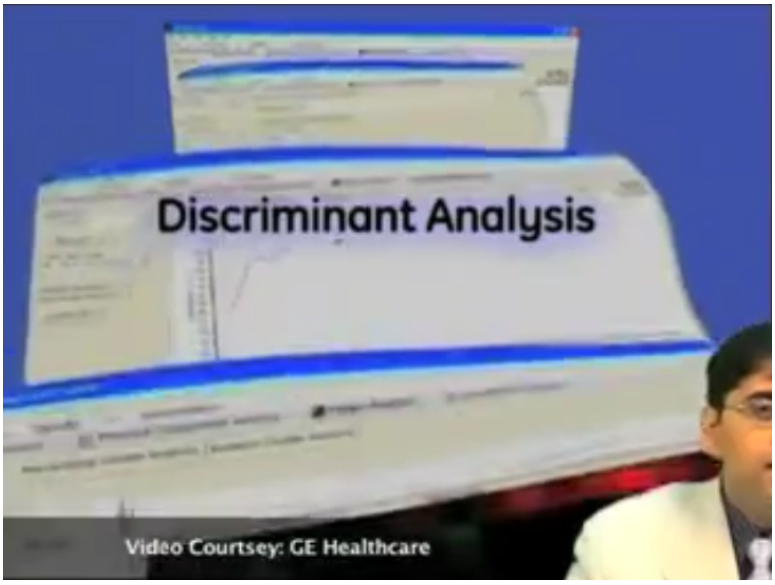
Fluorescence 2 dimensional, difference in gel electrophoresis or 2DE DIGE is an advanced 2DE technique that allows for accurate computation with statistical confidence while controlling the non biological variations.

(Refer Slide Time: 26:33)



In DIGE proteins extracted from different type of cells or tissue sample are labeled with different fluorescence agents such as (()) (26:47) mixed and then separated by 2 dimensional electrophoresis on a single gel. The proteins are detected separately using the excitation wavelength specific to the different fluorescent regions (()) (27:11).

(Refer Slide Time: 27:16)

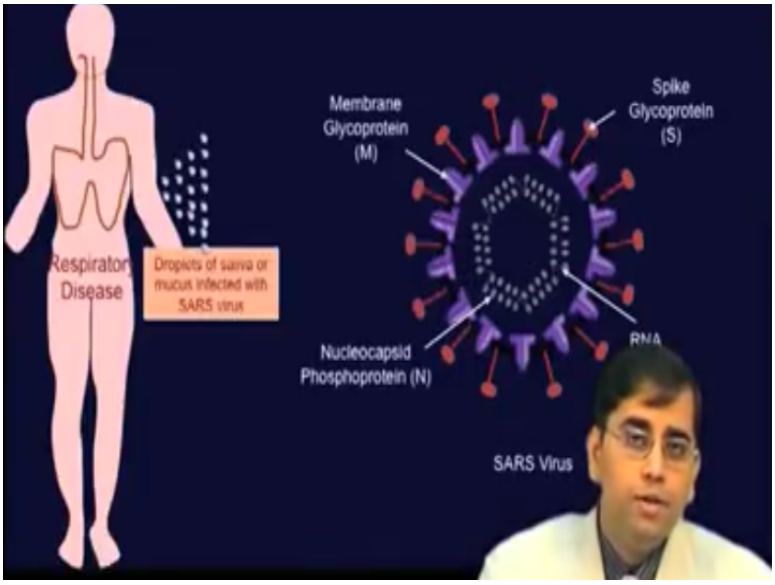


The commercial software such as (()) (27:17) facilitate the automated analysis of dye gels and provide differential expression analysis, principle component analysis, pattern and discriminate analysis.

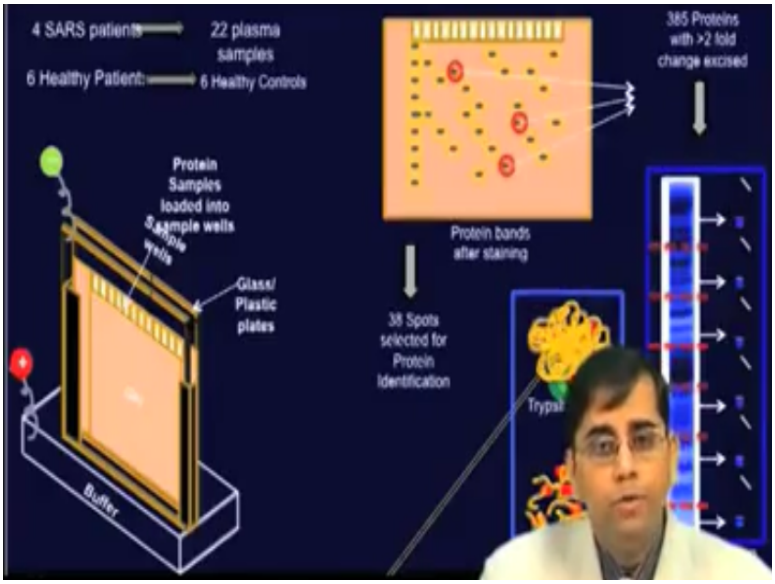
(Refer Slide Time: 27:35)

2 dimensional electrophoresis.

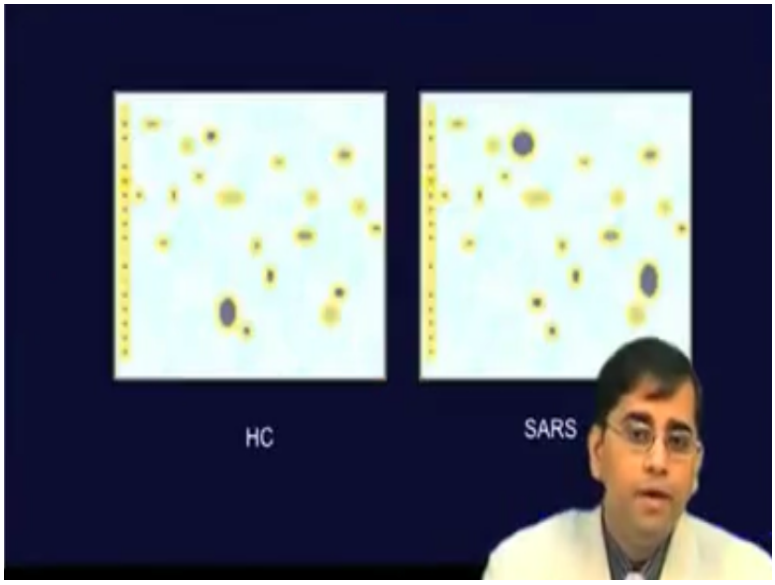
(Refer Slide Time: 27:37)



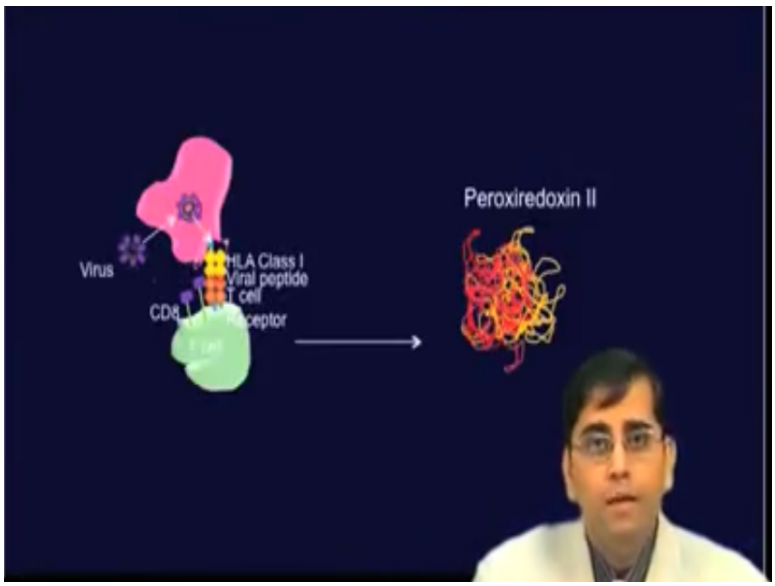
(Refer Slide Time: 27:40)



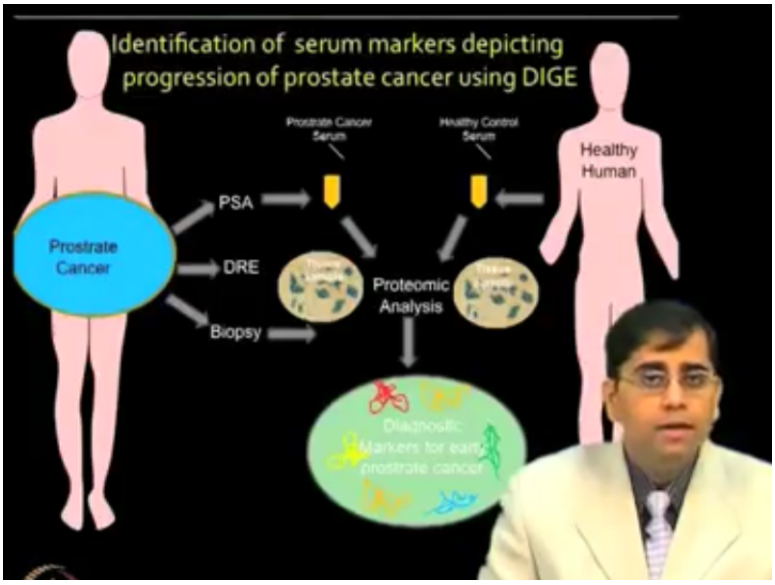
(Refer Slide Time: 27:43)



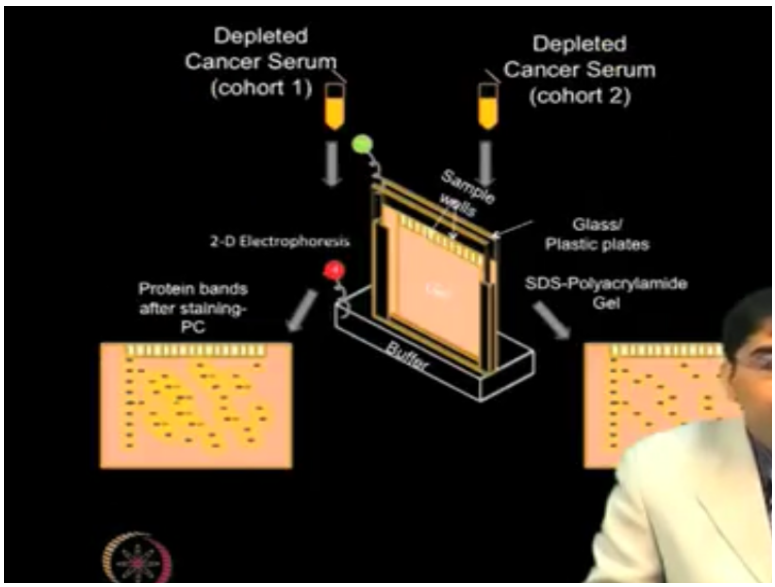
(Refer Slide Time: 27:45)



(Refer Slide Time: 27:47)

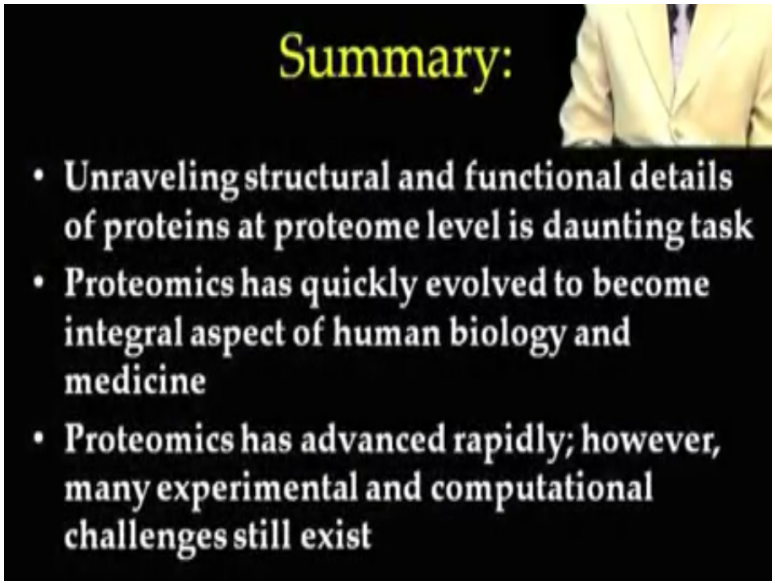


(Refer Slide Time: 27:48)



DIGE followed by mass spectrometric technique has been applied for many applications. Some of these applications will be discussed in this module of gel based proteomics. This century is considered as century of biology in which life science research is undergoing a profound transformation by employing various omic technologies.

(Refer Slide Time: 28:11)



In summary unraveling a structural and functional details of proteins at the proteomic scale is very daunting task. However, proteomics has come to mean virtually everything in protein research and it has quickly evolved to become an integral aspect of human biology and medicine. During the subsequent lectures I will take you to a journey of proteins and proteomic research by providing basic concepts and details of proteomic techniques.

I hope it will enthuse you to learn about proteomic techniques and proteomic concepts. Thank you for your attention.