Swayam Course - Analytical Techniques

Week: 10, Module 25 - Clinical Proteomics

Content Writer - Dr. Hari Prasad, Assistant professor, Department of Biophysics, All India Institute of Medical Sciences, New Delhi-110029

Learning Objectives:

- **Introduction**
- **Concept of Same Genome, Different proteome**
- **Need of clinical proteomics**
- **Clinical Proteomics: Platform for research**
- **Salient features of clinical proteomics and Rational of clinical proteomics**
- **Important aspects of clinical proteomics**
- **Comparative gel based proteomics and DIGE proteomics**
- **Validation Experiments**
- **Applications and Methodology**
- **Trypsin digest peptide mass finger printing**
- **Summary**

Introduction

A proteome is a collection of proteins of an organism's tissue at a particular time and under a set of physiological conditions. In the process of compilation of a proteome, proteins are identified and interactions between different set of protein or other bio-molecules is studied. Such interactions constitute different metabolic processes that occur in cells at all times. Proteomics is a comprehensive analysis of proteins in different cell types under various physiological conditions related to different functions, with changes in expression pattern being influenced by time, environmental changes, genetic factors and disease pathology. Proteomics is an extremely interesting field because a buildup of a complete proteome and study of all the above discussed goals has not been complete even for a single cell. Hence there is immense scope to unravel the complex world of proteins and each new study opens a new horizon waiting to be discovered. Researchers across the world are analysing sub-proteomes by selecting a set of

cells and conditions. This is done in order to understand the more complex picture of proteomics and the dynamics involved.

Concept of same genome and different proteome

Large-scale DNA sequencing has provided vast sequence data which can be studied for protein analysis. Data generated by mass spectrometry can be rapidly correlated by searching in databases for proteins and nucleotide sequences. Numerous pathways take place in a cell and there is enough evidence to indicate crosstalk between these pathways. Such advances as described above can be used to decipher the complex network of proteins involved in different biological processes by advanced protein expression studies and the identification of proteins in complexes. It should be noted that gene microarrays provide a snapshot of expression of many or all genes in the cell. However, mRNA levels may fluctuate and are not a reliable reflection of different types of proteins expressed in the cell. This may arise due to difference in stability of mRNA, non-representation of alternative splicing and variation in extent of translation. Further levels of protein may change due to proper folding/misfolding events and protein stability. Hence the study of proteomics is a more appropriate indicator of physiological, biochemical correlation of proteins in a cell that determine its functioning.

Clinical proteomics as a Platform for research

Clinical proteomics refers to the understanding of protein expression in the back drop various patho-physiological conditions that form the phenotype of healthy individuals and various clinical scenarios that a clinician attends to in a hospital. Management of a patient pertains to the clinician's ability to diagnose an ailment, differentiate it from other disease states and treat it with the appropriate medicine. As patient oriented research mainly involves diagnosis and therapeutics, clinical proteomics acts as an ideal foil for translational research wherein the end data is either a drug target or a biomarker. It is therefore not uncommon for problems to be taken from bed side to proteomics bench, where experiments are done to get solutions. These solutions are then taken back to the wards for the clinicians to use.

Salient features of clinical proteomics

- Protein expression profiling in various patho-physiological conditions
- Comparative proteomics of various phenotypes to delineate protein signatures that is very specific to a particular condition
- Associating protein regulation with clinical parameters

- Identification of biomarkers and drug targets
- Understanding patient response to pharmaco-therapy or chemotherapy
- Differentiating closely mimicking clinical scenarios
- Provides a good platform to understand disease biology
- Perfect interface for translational research
- Combining the proteome profile with patient examination and other investigations helps the clinician to gain better insights into the human body

Rational of clinical proteomics

Various environmental factors affect patho-physiological state of the body and, at the molecular level cause differential regulation of genes that alters protein expression Protein expression profiles can be mapped and differential expression between clinical phenotypes can be quantified 'Differentially expressed proteins' are then associated with 'Distinct human body states' with all other parameters being the same to the extent possible Unique protein signatures are proposed as being intricately related to causation of the disease and are therefore potential biomarkers or drug targets.

Important aspects of clinical proteomics

- Patient selection has to be stringent with specific inclusion and exclusion criteria
- Protein sample preparation (collection, processing, storage, experimental use) has to have well defined guidelines and protocol
- Choice of tissue is usually based on research Question, feasibility and patient compliance
- Presence of high abundant protein and ionic impurities need to be addressed
- While discovery is done using small set of samples; Validation is done using anti-body based assay on a large cohort of patient samples
- Gel based proteomics and non-gel based proteomics offer platforms of research with their own unique set of advantages, dis-advantages and challenges
- Sensitivity and accuracy of mass spectrometry is a factor in making the right identification

Applications of clinical proteomics

Clinical proteomics is an emerging interdisciplinary research field. It can be used for (i) early detection/diagnosis of disease, (ii) prediction of how a disease will behave over time and how a specific patient will respond to a given treatment, and (iii) identification of novel drug targets for therapeutic intervention. Identifying unique patterns of protein expression, or biomarkers, associated with specific

diseases is one of the most promising areas of clinical proteomics. One of the first protein biomarkers used in disease diagnosis was prostate-specific antigen, which is commonly used in diagnosing prostate cancer in men. Stable isotope labeling strategies in combination with mass spectrometry have been applied successfully to study the dynamics of modifications such as post translation modifications. The ultimate goal of proteomics is to characterize the information flow through protein networks. This information can be a cause, or a consequence of disease processes. Proteome profiles along with metabolomic profiling will revolutionize the understanding of cell function thereby providing a whole new dimension to the concept of human health.

Examples

- Biomarkers for differentiation of intestinal TB from Crohn's disease
- Biomarkers/Drug targets for Tubercular Meningitis
- Biomarkers for differentiation of advanced ovarian cancer
- Biomarkers for understanding chemo- response in advanced Ovarian cancer
- Biomarker's for Alzheimer's Disease
- Biomarkers for Leishmaniasis
- Biomarkers for Sentinel Lymph node metastasis in early Breast cancer
- Biomarkers to explain rational of co-administering stem cells with cytokines in Myocardial Infarction
- Biomarkers for monitoring pharmaco-therapy in Parkinson's disease and Schizophrenia

DIGE proteomics

Traditionally in 2D-PAGE, various samples are often separated in multiple gels. Those gels are then overlapped to compare one gel to another. However, due to differences in spatial resolution and spot intensities in different gels, the overlaying of images and correct matching of proteins is difficult. There can also be differences due to the following reasons: 1. Variation in protein uptake by the isoelectric focusing strips; 2. Incomplete protein transfer from the first to the second dimension gel; 3. Local inconsistencies in gel composition; 4. Field strength or pH gradients, and 5. Differential uptake of stain by proteins under different experimental conditions. These gel-to-gel variations can hide or interfere with the biological variations between the samples, particularly in low abundance proteins. DIGE is a technique that is developed to overcome the above problems. In DIGE, cyanine dyes are used to label the protein mixtures, prior to electrophoresis, which ensures comigration of proteins. Since this happens on the same

gel, the possibility of differences due to above discussed causes is eliminated. This is a technique used to analyze the differences in proteomic profile between cells/tissues in different functional states at particular time of growth. Firstly, samples are tagged with unique fluorescent dyes; secondly they are run together on the same 2D-PAGE gel. Different fluorescent images are generated and superimposed over each other. DIGE is employed for the study of proteins that are expressed differentially in various sources of samples. Technology is based on using cyanine dyes (Cy2, Cy3 and Cy5) used in DIGE are N-hydroxy succinimidyl ester derivatives that are covalently tagged to the _-amino group of lysine residue of proteins, and replaces the amino group positive charge with the positive charge of the dye. The binding of the dyes to the proteins leads to a homogenous increase in molecular weight of the all the proteins in the sample. These dyes are hydrophobic, therefore to prevent precipitation of proteins minimal labeling is performed i.e. only one lysine residue per protein is labeled. The dyes are all charge-matched and molecular mass-matched to prevent alterations of the isoelectric point and to minimize dye-induced shifting of labeled proteins during electrophoresis. There is also another class of cyanine dyes called saturation dyes, which saturate cysteine residues instead of minimally labeling lysine residues. These dyes are also mass and charge-matched. Saturation cysteine dyes are more sensitive than minimal lysine labeling. However, in the absence of cysteine residues, labeling of proteins is not possible causing it to remain undetectable. An internal standard should always be included in a DIGE experiment to enhance the normalization of spots across gels and reduce experimentation variation, increasing the accuracy of quantification, to help gel to gel comparisons and for better analysis of protein expression differences. DIGE is a better method to study differential protein expression because it reduces the need to run multiple gels and can be used to quantify proteins due to higher sensitivity and linearity as compared to other standard colorimetric staining methods.

Methodology in DIGE

- ☐ Samples (Experiment (diseased), control (normal) & Internal standard (pooled collection of all samples)) are labeled with Cy dye flours and are run on a single gel to avoid gel to gel variations and enable in gel analysis
- ☐ Rehydration;
- ☐ IEF;
- ☐ Equilibration
- ☐ SDS PSGE
- ☐ Labeled proteins are visualized using a Typhoon TRIO Variable Mode Imager (GE Healthcare, USA) by scanning Cy2, Cy3 and Cy5 images in different excitation wavelengths and emission filters.

- Gel images are processed using software that allows relative quantification of expression;
- consistency of expression pattern and normalization by co-detection image pairs, which intrinsically link a sample to its in-gel standard
- Preparative gels are run with higher amount of protein to enable staining and visualization of spots by naked eye to be picked for trypsin digestion and mass analysis

Trypsin digest peptide mass finger printing

The differentially expressed protein spots are excised and transferred to microfuge tubes. This is destained with 200µl of 40mM ammonium bicarbonate and 50% acetonitrile for 45 minutes. The gel pieces are then rehydrated for 5 minutes at room temperature in 100µlof 100% acetonitrile. The acetonitrile is removed completely. The gel pieces are dried in speed vacuum for 10 to 15 minutes. Trypsin, an enzyme that is site specific, is added 500ng per protein spot with 20µl digestion buffer to re-swell the gel pieces. This is incubated on ice for 1 hour. The digestion buffer is then added to completely cover the gel and the tubes are incubated at 37˚C overnight. The liquid is transferred to a fresh tube and 50µl of MiliQ water is added to the gel containing tube and mixed by tapping for 10 minutes. The liquid is transferred to the tube. The protein is extracted by incubating the gels with extraction buffer (acetonitrile 5ml, formic acid 0.5ml, water 4.5ml) at room temperature for 30 minutes, two times. The tubes with the extraction buffer are centrifuged for 5 minutes. The supernatant is collected and lyophilized. The lyophilized powder contains the peptides that can be subjected to nano-Liquid chromatography and MS analysis for identification.

Nano-Liquid chromatography

Lyophilized peptides are reconstituted in 20µl of 50% acetonitrile and 0.1% formic acid solution for injection into a nano-LC system. Peptide mixture are separated using a 70 minute linear gradient from 5% to 98% acetonitrile in 0.1% formic acid with a flow rate of 300 nl/min by reverse phase chromatography using a Pep Map C18 column before sequential elution based on hydrophobic index of the peptides. Eluted peptides that come out through a small orifice at high electric voltage are converted charged gaseous molecules called ions that enter mass spectrometry machine.

Mass spectrometric analysis

Mass spectrometers are composed of three essential parts. First is an ionization source which converts molecules into gas-phase ions. Once ions are created, individual mass-to-charge ratios (m/z) are separated

by a second device known as a mass analyzer and transferred to the third, an ion detector. A mass analyzer uses a physical property [e.g. electric or magnetic fields, or time-of-flight (TOF)] to separate ions of a particular m/z value that then strike the ion detector. The m/z value of the ion is determined by the magnitude of the current that is produced at the detector as a function of time. Development of MALDI and ESI has been very important for proteomic studies.

Matrix-assisted laser desorption ionization (MALDI) produces ions by causing excitation of molecules that are isolated from the energy of the laser by an energy absorbing matrix. The laser energy then strikes the crystalline matrix that leads to excitation of the matrix followed by ejection of matrix and analyte ions into the gas phase. **Electrospray ionization (ESI)** creates ions through application of a potential to a flowing liquid which leads to development of charge in it producing an electrospray. The latter creates minute droplets of solvent-containing analyte from which the solvent is removed as the droplets enter the mass spectrometer by heat or some other form of energy (e.g. energetic collisions with a gas). This creates multiply-charged ions in the process. These ionization techniques have led to many developments in design and application of mass spectrometers. Different types of information can be generated like the accurate determination of molecular weight.

Information produced by tandem mass spectrometers (MS/MS), can be used to determine amino acid sequence. The progress of each run is monitored by recording the total ion current (TIC) for positive ions as a function of time in the m/z range of 400–1600 for MS and 140–1600 for MS/MS. The spectrum is acquired in an information dependent manner utilizing the Analyst QS 2.0 software 2.0 acquisition features to generate raw data in the *.wiff format without merging putatively like spectra. The other parameters that are set are: interface temperature, 50°C; curtain gas flow, 1.13L/min; declustering potential, 60V; focusing potential, 280V; declustering potential 2, 15V.

Protein identification

MS/MS analysis and data base searching was performed using either *Protein PilotTM* or *MASCOTTM* software version 2.0. All *Protein PilotTM* soft ware data base searches are performed using *ParagonTM* database algorithm for searching in through mode. Modifications considered are oxidation of methionine and carbamidomethylation of cysteine while searching. Search is further refined to include peptides with charged state from +2 to +3 and limited to *Homo sapien* species. The peptide mass tolerance range is ± 1.2 kDa and fragment mass tolerance is ± 0.6 kDa. The confidence limit for the protein identification is set at a minimum of 95%.

Validation experiments

The experiments explained above is for the discovery phase of the differentially expressed proteins on a certain number biological samples in each arm of the experiment for a possible analysis.

Some of the salient features of validation experiments are:

- Population based study on large cohort to off-set individual variations
- Number of samples depending on the incidence of disease; patients registered, patients screened, statistical justification and logistic/ethical feasibility
- Comprises of Anti-body based assays like ELISA, Western blot, Immuno-histochemistry
- Mass spectrometry based MRM analysis and RT-PCR
- Helps to derive statistical parameters such as sensitivity, specificity, positive predictive value, which reflect the quality of biomarker as a diagnostic tool
- Eliminate systemic mass measurements in LC-MS/MS based experiments due to factors like: hydrophobicity, ionization, mass differentiation of L-I and K-Q
- Turn around time is fast and antibody/primer based experiments are accurate

Applications with examples

☐ **Proteome analysis of the macroscopically affected colonic mucosa of Crohn's disease and intestinal tuberculosis to identify biomarkers that can differentiate the two**

Differentiation between intestinal tuberculosis (ITB) and Crohn's disease (CD) is challenging in geographical regions where both these diseases are prevalent. There is a need of biomarkers for differentiation between these two disorders. Colonic biopsies from inflamed mucosa of treatment naive patients with ITB, CD and controls were used for analysis. Protein extracted from biopsies was digested with trypsin and resulting peptides were labeled with iTRAQ reagents. The peptides were subsequently analyzed using LC-MS/MS for identification and quantification. Gene ontology annotation for proteins was analyzed in PANTHER. Validation experiments were done for six differentially expressed proteins using immunohistochemistry. 533 proteins were identified and 241 proteins were quantified from 5 sets of iTRAQ experiments. While 63 were differentially expressed in colonic mucosa of patients with CD and ITB in at least one set of iTRAQ experiment, 11 proteins were differentially expressed in more than one set of experiments. Six proteins used for validation using immunohistochemistry in a larger cohort of patients; none of them however was differentially expressed in patients with ITB and CD. There are

differentially expressed proteins in tissue proteome of CD and ITB. Further experiments are required using a larger cohort of homogeneous tissue samples.

- **Two-dimensional difference gel electrophoresis (DIGE) analysis of sera from visceral leishmaniasis patients for biomarker discovery**

Visceral leishmaniasis is a parasitic infection caused by *Lesihmania donovani* complex and transmitted by the bite of the phlebotomine sand fly. This study was carried out to discover differentially expressed proteins which could be potential biomarkers. Sera from six visceral leishmaniasis patients and six healthy controls were depleted of high abundant proteins by immunodepletion. The depleted sera were compared by 2-D Difference in gel electrophoresis (DIGE). Differentially expressed proteins were identified the by tandem mass spectrometry. Three of the identified proteins were further validated by western blotting. Proteins alpha-1-acidglycoprotein and C1 inhibitor were up-regulated and transthyretin, retinol binding protein and apolipoprotein A-I were down regulated in visceral leishmaniasis sera in comparison with healthy controls. Western blot validation of C1inhibitor, transthyretin and apolipoprotein A-I in a larger cohort ($n = 29$) confirmed significant difference in the expression levels ($p < 0.05$). The five proteins identified here have potential, either independently or in combination, as prognostic biomarkers.

- **Apolipoprotein A1 as a potential biomarker in the ascitic fluid for the differentiation of primary from metastatic advanced ovarian cancers**

Primary ovarian cancer and ovarian metastasis from non-ovarian cancers in advanced stage are closely mimicking conditions whose therapeutics and prognosis are different. The objective was to therefore identify biomarkers that can differentiate the two variants of advanced ovarian cancers. Gel-based proteomics and antibody-based assays were used to study the differentially expressed proteins in the ascitic fluid of fourteen patients with advanced ovarian cancers. Programmed Cell Death 1-Ligand 2, apolipoprotein A1, apolipoprotein A4 and antihuman fas antibody were differentially expressed proteins.

Apolipoprotein A1 with a 61.8_ng/ml cut-off is a potential biomarker with the best differentiating statistical parameters.

- **Comparative Proteomic Analysis of Advanced Ovarian Cancer Tissue to Identify Potential Biomarkers of Responders and Nonresponders to First-Line Chemotherapy of Carboplatin and Paclitaxel.**

Conventional treatment for advanced ovarian cancer is an initial debulking surgery followed by chemotherapy combination of carboplatin and paclitaxel. Despite initial high response, three-fourths of these women experience disease recurrence with a dismal prognosis. Patients with advanced-stage ovarian cancer who underwent cytoreductive surgery were enrolled and tissue samples were collected. Post surgery, these patients were started on chemotherapy and followed up till the end of the cycle. Fluorescence-based differential in-gel expression coupled with mass spectrometric analysis was used for discovery phase of experiments, and real-time polymerase chain reaction, Western blotting, and pathway analysis were performed for expression and functional validation of differentially expressed proteins. While aldehyde reductase, hnRNP, cyclophilin A, heat shock protein-27, and actin are upregulated in responders, prohibitin, enoyl-coA hydratase, peroxiredoxin, and fibrin are upregulated in the nonresponders. The expressions of some of these proteins correlated with increased apoptotic activity in responders and decreased apoptotic activity in nonresponders. Therefore, the proteins qualify as potential biomarkers to predict chemotherapy response.

- **Apolipoproteins and complement factors as biomarkers of Alzheimer's disease**

Alzheimer's disease is the most common cause of dementia in elderly persons. Quick diagnosis of Alzheimer's disease will allow treatments that may help slow its progression. The correlation between cerebrospinal fluid (CSF) parameters and progression of Alzheimer's disease is higher than and independent of other risk factors. We have compared sixteen CSF samples of clinically diagnosedAlzheimer's disease patients with non demented subjects using proteomics approach. Apolipoprotein E, apolipoprotein J, complement C4b, hemopexin and complement factor B were identified as differentially expressed proteins. Pathway analyses show that these proteins have interacting partners in Alzheimer's and apoptotic pathways.

- **Two dimensional difference gel electrophoresis analysis of cerebrospinal fluid in tuberculous meningitis patients.**

Tuberculous meningitis (TBM) is a serious complication of tuberculosis that affects the central nervous system. Present methods to diagnose TBM are not suitable for early diagnosis. Molecular markers and sensitive methods to identify them in the early stage of infection of TBM are critically needed for efficient management. We have done the proteomic analysis of TBM cerebrospinal fluid (n=20) with 2-dimensional difference gel electrophoresis (2D-DIGE) and mass spectrometry. 11 human proteins and 8 mycobacterial proteins with changed expression levels were identified in comparison to controls.

Arachidonate 5-lipoxygenase and glial fibrillary acidic protein, two of the identified proteins, were validated with western blot technique on a larger set of disease and control samples (n=40). These two proteins were also analyzed in fungal meningitis samples for the sake of speficity. Arachidonate 5-lipoxygenase can be considered for validation as a potential marker for diagnosis of TBM.

Summary

- **Clinical proteomics pertains to study of protein expression in human body**
- **Delineation of specific protein signatures in patho-physiological conditions is useful to understand molecular events in a disease**
- **Applications of clinical proteomics are: Understanding biology of the disease;**
- **Biomarker discovery; Drug target identification; knowing drug respose, and differentiating closely mimicking clinical states**
- **Gel based comparative proteomics and Non-gel based proteomics are used in clinical proteomics**
- **Advancements in mass spectrometry- sensitivity and accuracy has propelled the pace of work in clinical proteomics**
- **Scope for clinical proteomics to develop personalized medicine**