LECTURE-34

Serum Proteomics

Handout

PREAMBLE

Serum is one of the attractive biological samples for the clinical studies because of its ease of collection, less invasive nature and it shows the proteomic alterations in response to the diseases. Serum is one of the complex mixtures containing large number of proteins with dynamic ranges of concentrations in the order of 10. Presence of large number of interfering agents, patient to patient variation and lack of reproducibility makes the serum proteomic study for the discovery of biomarkers challenging. Various gel-based and mass spectrometry based proteomics methods have been employed to study serum and plasma proteomics. Although major advancements have happened in the field of proteomics, but due to the challenges of complex proteome, till now only few biomarkers are translated for clinical applications.

PROF. SANJEEVA SRIVASTAVA
DEPARTMENT OF BIOSCIENCES & BIOENGINEERING
IIT BOMBAY, MUMBAI, INDIA

OUTLINE OF LECTURE

I. Blood, serum and plasma proteome

II. Serum proteomics for biomarker discovery

(a) Serum biomarkers in cancer

(b) Biomarker discovery in autoimmune diseases

(c) Biomarker discovery in infectious diseases

(d) Biomarker discovery in heart diseases

III. Challenges for the serum proteome analysis

IV. Recent advancements in serum proteomics

V. HUPO Human Plasma Proteome Project (HPPP)

VI. Conclusions

I. Blood, serum and plasma proteome

Blood is a fluid connective tissue containing different types of cells in the fluid called plasma. The blood cells consist of erythrocytes, leucocytes and blood platelets. Erythrocytes play major role in the transport of oxygen and carbon dioxide from lungs to various parts of the body and vice versa. Leucocytes play major role in providing immunity to the body from the foreign antigens (pathogens). Platelets help in the blood coagulation. The extracellular fluid matrix part of the blood is called as plasma, which consists of various kinds of proteins, carbohydrates, lipids, blood coagulation factors and salts. Proteins play major role in transport of lipids, hormones, peptides and many other molecules.

Serum is plasma without fibrinogen. Serum contains large number of proteins, where the difference between the high abundant proteins with respect to the low abundant proteins is around 10 orders of magnitude. All the serum proteins are either secreted/released from various tissues, blood cells or damaged tissues. In disease conditions the increase/decrease in the serum levels of proteins is observed because of over-expression/ under-expression of certain genes in cells in various tissues. In disease condition we may find new proteins in serum because of mutation in certain genes or released from the pathogens. These proteins can be useful in diagnosis of the disease. Proteins, which could differentiate the disease and healthy conditions, are called as diagnostic biomarkers (Ray et al., 2011).

II. Serum proteomics for biomarker discovery

Because of the ease of sample collection serum is one of the attractive sources for the discovery of biomarkers for the diagnosis and prognosis of various diseases. With the advances in the field of proteomics various sensitive techniques are developed which can detect the proteins in the nanomolar concentrations.

Illustration: Serum proteome analysis for biomarker discovery

Serum proteome analysis for biomarker discovery is illustrated in following animation. The gel-based proteomic analysis involves protein extraction, separation, gel analysis followed by protein identification using mass spectrometry.

(a) Serum biomarkers in cancer

Cancer is a diseased state where the modified cells undergo rapid uncontrolled proliferation to give unwanted mass of cells called tumor tissue. These tumor tissues invade into the normal tissue regions and damage the surrounding tissues. In cancer conditions large number of mutations takes place at the genomic level, which results in the proteomic alterations in that organism. The tumor tissue proteins get leaked into the blood and can be further identified by using various high throughput proteomic techniques. Regardless of which disease is under investigation and what technique is used, depletion of abundant proteins becomes very crucial (Fig 1).

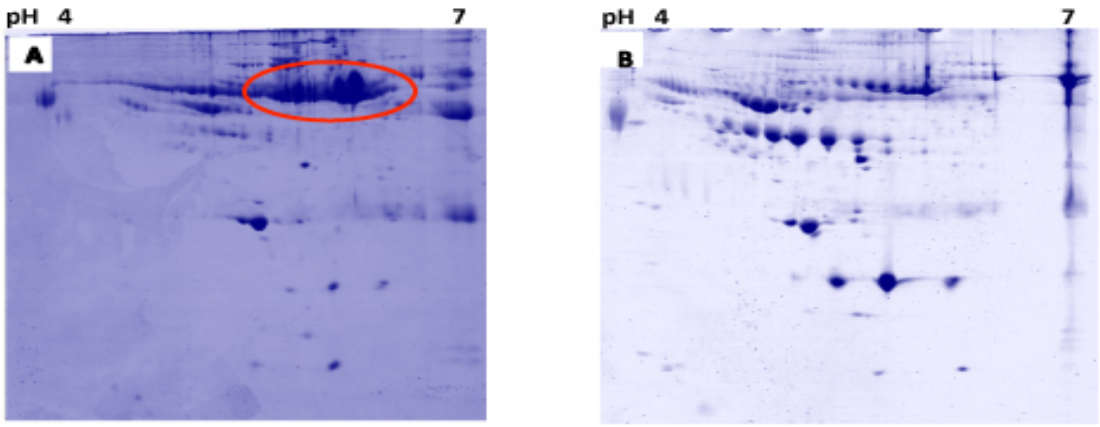

Fig 1. 2DE gels representing the serum proteome A) crude serum before depletion, and B) After depletion of abundant proteins.

Various proteomic techniques have been used for the identification of biomarkers in various cancers like ovarian, lung, pancreatic, brain cancer etc., (Kumar et al., 2010; Sun et al., 2007; Patz et al., 2007; Zhang et al., 2004). Schematic of two dimensional gel electrophoresis and mass spectrometry is shown in Fig 2 and Fig 3. For the clinical diagnosis purpose it is best to have single biomarker, which is specific for a disease. But in case of cancers because of their large heterogeneity it is not possible to have a

single biomarker for the diagnosis of the disease, but it is possible to have a panel of candidate proteins showing some unique pattern in their expression levels can be used as biomarkers. Even though such a panel of biomarkers is not cost effective for clinical diagnosis, but the accuracy for the disease diagnosis will be more when compared to the single biomarker system.

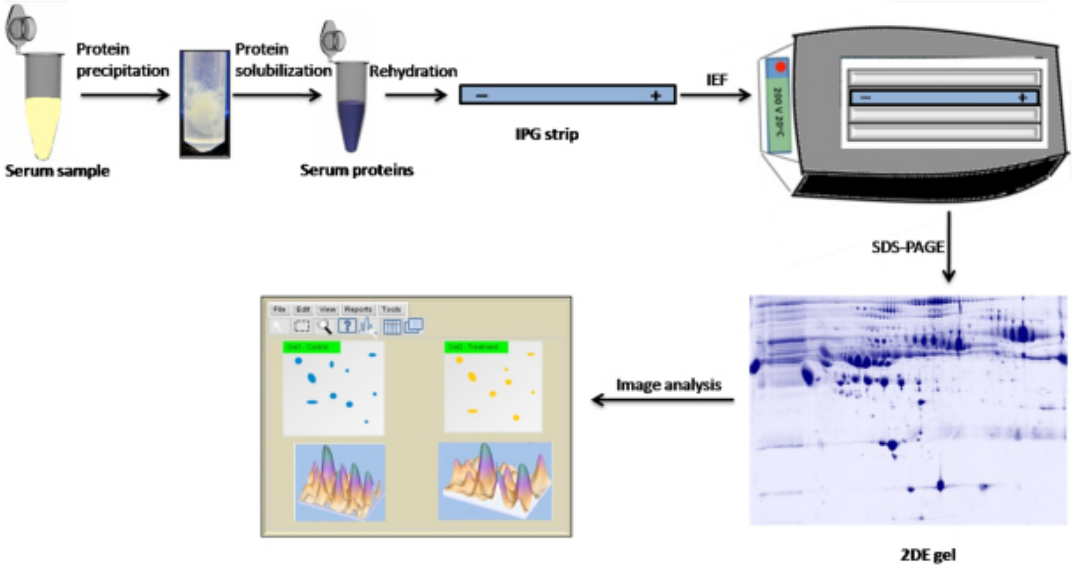

Fig 2. Schematic representation of Two dimensional gel electrophoresis for the serum proteome analysis.

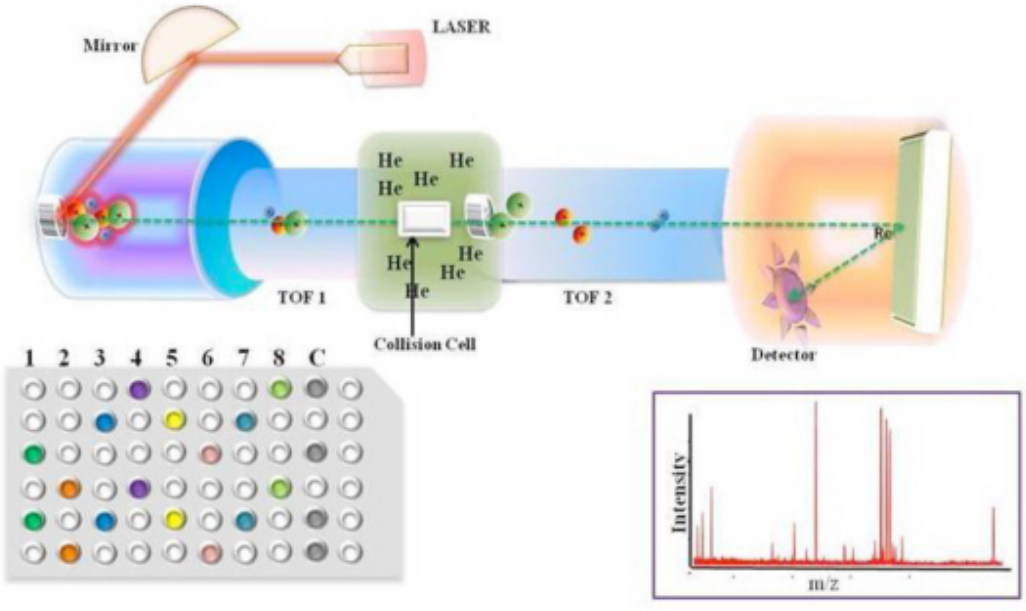

Fig 3. Schematic representation of MALDI-TOF/TOF for the detection of peptide mass and protein identification.

(b) Biomarker discovery in autoimmune diseases

Auto-immune diseases are those in which our body generates antibodies against self antigens. This is because of the malfunctioning of our immune system. Auto-antibodies form immune complexes by interacting with the specific antigen present on the cells of an organ and results in the deterioration of the tissue, necrosis and inflammation. Auto-antibodies are found in various autoimmune disorders like multiple sclerosis, Rheumatoid arthritis and inflammatory bowel disease etc. (Liao et al., 2004, Rithidech et al., 2009, Meuwis et al., 2007). Auto-antibodies can be found in various cancers which are generated against the altered proteins resulting from the mutations in the corresponding genes. These auto-antibodies could be useful in the early diagnosis of the disease.

(c) Biomarker discovery in infectious diseases

The diseases which are caused by external pathogenic agents like bacteria and other micro-organisms and transmitted from one person to other directly or indirectly by means of vectors are called as infectious diseases. These pathogens produce their own proteins for their survival and from being escaped from the host immune system. Our body also responds to the infection by showing altered expression of proteins and changing the physiological conditions like increasing the body temperature. The pathogen proteins or the alteration in specific serum protein levels can be used as a diagnostic marker for the disease. Proteomic techniques are used to identify the serum biomarkers for various infectious diseases like chronic hepatitis-B, malaria, severe

acute respiratory syndrome, dengue hemorrhagic fever etc. (He et al., 2003; Kassa et al., 2011; Ren et al., 2004; Thayan et al., 2009).

(d) Biomarker discovery in heart diseases

According to WHO, the heart diseases (Cardiovascular diseases) are dramatically increasing all around the world. The cardiovascular diseases (CVD) are responsible for 30% of deaths across the world. Cardiovascular diseases includes a group of complications like coronary heart disease, hypertension, heart failure, congenital heart diseases etc., Norepinephrine, endothelin, atrial natriuretic peptide (ANP), B-type natriuretic peptide (BNP) are some of the candidate proteins found to be altered in cardiovascular diseases. Serum angiopoietin-2 is one of the serum makers, which is showed to be elevated with the advancement in the cardiovascular diseases. Serum biomarkers for the cardiovascular diseases are useful in the early detection of disease and reduce the deaths because of CVD. The serum biomarkers have been identified in various diseases (Jarolim. 2006) and are described in Table 1.

Table 1: Serum biomarkes in various diseases

Disease	Proteomic approach	Serum biomarkers identified
A) Cancers		
Gliomas	Western blotting, 2-DE, MALDI-TOF MS	Haptoglobin α2 chain, Apolipoprotein A-I, Angiopoietin 2 isoform b precursor (Kumar et al., 2010)
Pancreatic cancer	2-DE, MALDI-TOF MS	Cyclin I and GDI2 (Rab GDP dissociation inhibitor β (Sun et al., 2007)
Lung cancer	2D-DIGE, MALDI-TOF MS	Carcinoembryonic antigen, squamous cell carcinoma antigen, α1-antitrypsin and retinol binding protein (Patz et al., 2007)
Ovarian cancer	SELDI-TOF MS	Apolipoprotein A1, truncated form of transthyretin and cleavage fragment of inter-α-trypsin inhibitor heavy chain H4 (Zhang et al., 2004)
B) Autoimmune diseases		
Rheumatoid arthritis	2D-LC-MS/MS	CRP, calgranulin A, calgranulin B and calgranulin C (Liao et al., 2004)
Multiple sclerosis	2-DE, MS	Apoliprotein-C-III, serum amyloid P component, complement factor-I, clusterin, gelsolin, hemopexin, kininogen-1, hCG1993037-isoform, and vitamin D-binding protein (Rithidech et al., 2009)
Inflammatory bowel diseases	SELDI-TOF MS	PF4, MRP8, FIBA and Hpα2 (Meuwis et al., 2007)
c) Infectious diseases		
Hepatitis B	2-DE, MALDI-TOF MS	Haptoglobin β and α2 chain, Apolipoprotein A-I and A-IV, alpha1-antitrypsin, Transthyretin and DNA topoisomerase II β (He et al., 2003)
Dengue hemorrhagic fever	2-DE, MALDI-TOF MS, ELISA, Westernblotting,	NS1 and α1-antitrypsin (Thayan et al., 2009)
Severe acute respiratory syndrome	2-DE, 2-D Western blotting	Truncated forms of a1-antitrypsin (Ren et al., 2004)
Malaria	SDS-PAGE,	Serum amyloid A, C reactive protein,

	Western blotting, ELISA	Apolipoprotein E, LPS binding protein, complement components C1q, C4, C3 and C5a (Kassa et al., 2011)
d) Heart diseases		
Heart failure	ELISA	Serum soluble ST2 (Weinberg et al., 2003)

III. Challenges for the serum proteome analysis

Despite the advancements in the field of proteomics still there are challenges to study serum proteome for the identification of disease specific biomarkers. The challenges include pre-analytical, technical and biological issues. Pre-analytical issues include the serum sample collection, storage and handling the clinical sample. The technical issues include the reproducibility of the data, sensitivity of the technique and the biological issues include the complexity of the serum sample and the patient-to-patient variations etc. (subject selection) (Ebert et al., 2006).

Pre-analytical issues:

Variations in serum proteomic experimental results are mainly due to variations in sample collection, handling and storage processes. Usage of collection tubes made of different material (some materials adsorb the proteins on their surface), storage temperature of serum and protein samples are some of the pre-analytical reasons for the irreproducible data. The process of serum sample extraction from the blood and the experimental conditions like centrifugal force (g) and the processing temperature and number of freeze thaw cycles are factors which have been shown to affect the results. In order to prevent such kind of variations, one needs to use same kind of serum separation tubes and maintain same experimental and storage temperature conditions for the serum samples. To avoid large number of freeze thaw cycles serum samples should be stored in the form of small aliquots.

Technical issues:

These include the sensitivity of technique used for the analysis of the clinical samples. Each technique is having it's own sensitivity. For example sensitivity of the CBB dye used in the classical two-dimensional gel electrophoresis (2DE) method is ~2ng (Using modified Kang's CBB staining protocol). The low abundant proteins having their concentrations in the order of 10 can't be detected by using such classical 2DE method. Identification of the proteins by using mass spectrometry also requires either pre-fractionation of the protein samples or separation of the proteins on the gel. Gel based proteomic techniques are limited with their detection methods/ MS compatibility issues. LC-MS/MS is one promising technique for the identification of biomarkers. As these experiments are very expensive, experimental cost is another limitation.

Biological issues/ Sample issues:

Serum sample complexity:

Serum is a very complex mixture. It contains large number of proteins with dynamic range of concentrations in the order of 10. With the limitations in the current techniques we can't detect such low abundant proteins in the limited amount of clinical sample.

High abundant proteins:

Serum contains high abundant proteins like albumin and immunoglobulins. Albumin alone accounts for 50% of total serum proteins. There are total 22 high abundant proteins are there in the serum, which masks the detection of low abundant proteins. Most of the biomarkers are expected as low abundant proteins. Therefore, prior to the serum proteomic study it is essential to deplete the high abundant proteins in order to

increase the effective concentration of the low abundant proteins in the serum protein sample to be analyzed.

Post-translational modifications:

Proteins undergo large number of post-translational modifications (PTMs) under normal and pathological conditions. These PTMs may interfere with the detection of proteins.

Interfering agents:

Serum contains large number of chemical species like proteins, salts, carbohydrates, fatty acids, cholesterol, hormones etc. So it is challenging to study the proteomic alterations in such kind of crude samples. Presence of any kind of interfering agents in the protein sample may interfere with the process of protein separation/ identification. For example, high amount of salts and detergents in the protein sample interferes with the isoelectric focusing in the two-dimensional gel electrophoresis. In the same way high amount of salts interferes in the identification of proteins by using mass spectrometry. Therefore, prior to the analysis of the serum samples all the interfering agents should be removed from the serum proteome sample by using different precipitation and desalting methods.

Patient-to-patient variability:

Selection of patients is one of the important aspects in designing the experiment. Age and sex matched patients who haven't undergone any kind of prior treatment should be selected for the study. However the genetic, environmental and dietary variations reflect in the results, which are inevitable in the present scenario. Control subjects suffering

from other diseases/ undergoing any medication or having the habits like smoking, alcoholism should be eliminated from the study.

IV. Recent advancements in serum proteomics

One of the major challenges in serum proteomics is the presence of high abundant proteins like albumin and immunoglobulins in the serum. With advancement in filtration and chromatography techniques now it is possible to deplete the high abundant proteins from the serum. Using centrifugal ultra-filtration (molecular weight cut off filters) method the high abundant proteins can be removed and low abundant proteins can be concentrated. However, in this method high molecular weight protein, which are having molecular weight greater than the high abundant proteins are also lost. The solid phase extractions include antibody-based, metal ion based and dye binding methods. Among all of these solid phase extraction methods, antibody-based methods for the depletion of high abundant proteins are more promising method because of its affinity towards the abundant proteins. Now commercially large numbers of depletion columns are available for depleting 2 (albumin and IgG), 6 and 20 high abundant proteins from the serum. Depletion of these high-abundant proteins increases the effective concentration of the low abundant proteins, which are expected as the important biomarkers for most of the diseases. Even though antibodies are specific for the proteins against which they are raised, still some cross reactivity remains, because of which they bind to some other serum proteins also (Zhou et al).

Classical 2DE method can separate hundreds of proteins on the gel whose isoelectric points are in the range of 3-10. This classical 2DE method is having limitations like gel-to-gel variations, spot detection sensitivity and reproducibility of the gels. However, 2D-DIGE technique is an advanced form to overcome the limitations of the classical 2DE method. The fluorescent dyes (CyDyes) increases the sensitivity of

the technique and test and control samples can be run on the same gel, which eliminates the gel-to-gel variations. Still gel-based proteomic techniques have some limitations like they can't detect the proteins with pI in the more acidic, basic regions and the hydrophobic proteins like the integral membrane proteins. In order to overcome these limitations, various chromatography techniques are developed for fractionating the protein samples based on their physico-chemical properties. These fractions are further subjected to mass spectrometry for the identification of proteins.

The Isobaric tags for relative and absolute quantitation (iTRAQ) is one of the recent advancements in the clinical proteomic studies. This method is useful in multiplexing large number of samples at a time, which reduces the experimental errors (handling errors) and provide reproducible data. iTRAQ reagents contains molecular mass tags which are having a mass difference of 1 unit. These iTRAQ reagents are used to label different in-solution digested protein samples followed by separation of these labeled peptides by using different liquid chromatography techniques (strong cation exchange chromatography and reverse phase chromatography). The peptide fractions are subjected to MS/MS, where the mass tags are separated and the relative quantities of those particular proteins are measured. By using iTRAQ method we can label 4 or 8 samples at a time (4-plexing or 8-plexing).

Label-free proteomic techniques are gradually replacing the label-based techniques because of their sensitivity, low cost and less laborious nature. SELDI-TOF MS (Surface enhanced laser desorption and ionization-Time of flight mass spectrometry) is one of the promising proteomic techniques for the identification of the proteins having specific physicochemical properties. The SELDI platform includes a

surface which is having modified chemistry for binding the proteins of interest (having specific physicochemical properties). After sample application washing should be done for the removal of unbound proteins. The proteins of interest bind to the surface of SELDI plate, which enhances their effective concentration. Various kinds of surface modifications are available for detecting the hydrophobic proteins (integral membrane proteins), glycoproteins and phosphoproteins etc. This chip-based approach requires less amount of clinical samples, which is one of the advantages of this technique (Tang et al., 2004). Surface Plasmon resonance (SPR), Surface Plasmon resonance imaging (SPRi), nanowires and tubes, Chip based nano-LC-MS methods and microcantilevers are some of the emerging label free proteomic techniques. SPR works on the principle of changes in the refractive index (Ladd et al., 2009). This technique is widely used for interaction studies like the study of protein-protein, Receptor-ligand binding, protein-drug binging studies etc. Nanowires and tubes use the changes in the electrical conductance as its principle for the detection of molecular interactions (Chua et al., 2009). These label free techniques have advanced the proteomics field for the discovery of biomarkers from complex clinical samples like serum.

V. HUPO Human Plasma Proteome Project (HPPP):

Human proteome organization (HUPO) was formed in 2001 with aim of accelerating developments in the field of proteomics by collaborating with the major research institutes at the international level. HUPO started Human Plasma Proteome Project (HPPP) with following objectives: (1) to study the human plasma and serum proteome, (2) to identify the biological factors responsible for the variations within the individuals with respect to the time period, (3) to study the level of variations among the people in population because of their genetic make-up and their habits.

First HPPP meeting was held in Sept. 2002, in Ann Arbor. First time the HPPP data was presented in the 2nd HUPO World Congress in Montreal in Nov 2003. All the data obtained from the collaborators all around the world is stored, analyzed and findings are shared. This ambitious project will be helpful to understand and overcome various issues related with the complex serum/ plasma samples (Omenn et al., 2005).

VI. Conclusions

Serum is one of the complex biological samples having dynamic range of proteins. The ease of serum sample collection makes it as an attractive source for the detection of biomarkers. Serum contains high abundant proteins, which make the detection of low abundant proteins very tedious process. The advancement in the instrumentation of the mass spectrometry resulted in the development of very sensitive mass spectrometers, which enabled us to detect even the very low abundant molecules from the given sample. Label based, multiplexing methods were developed to study the quantitative changes in the clinical samples. Later on development of label-free detection methods advanced due to their sensitivity, less laborious and less economical nature. Large numbers of research institutes all around the world are working on the serum proteomics for the detection of biomarkers for various diseases; however, further efforts are required to transform the biomarker discovery from research level to the clinical application level.

References

- Ray, S., Reddy, P. J., Jain, R., Gollapalli, K. et al., Proteomic technologies for the identification of disease biomarkers in serum: advances and challenges ahead. Proteomics 2011, 11, 2139-2161.
- Jarolim, P., Serum biomarkers for heart failure. Cardiovasc. Pathol. 2006, 15:144-149.
- Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W. et al., Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 2005, 5(13):3226-45.
- Tang, N., Tornatore, P., Weinberger, S. R., Current developments in SELDI affinity technology. Mass Spectrom Rev. 2004, 23, 34-44.
- Seibert, V., Ebert, M. P., Buschmann, T., Advances in clinical cancer proteomics: SELDI-ToF-mass spectrometry and biomarker discovery. Brief Funct. Genomic Proteomic. 2005, 4, 16-26.
- Ebert, M. P., Korc, M., Malfertheiner, P., Röcken, C., Advances, challenges, and limitations in serum-proteome-based cancer diagnosis. J. Proteome Res. 2006, 5, 19-25.
- Wang, X., Yong, H., Mi, L., Bai, Y. et al., Changes and significance of serum angiopoietin-2 levels in patients with coronary heart disease. Biomarkers. 2012, 17, 745-749.
- Kumar, D. M., Thota, B., Shinde, S. V., Prasanna, K. V. et al., Proteomic identification of haptoglobin α2 as a glioblastoma serum biomarker: implications in cancer cell migration and tumor growth. J. Proteome Res. 2010, 9, 5557-5567.
- Sun, Z. L., Zhu, Y., Wang, F. Q., Chen, R. et al., Serum proteomic-based analysis of pancreatic carcinoma for the identification of potential cancer biomarkers. Biochim. Biophys. Acta. 2007 1774, 764-771.
- Patz, E. F., Jr. Campa, M. J., Gottlin, E. B., Kusmartseva, I. et al., Panel of serum biomarkers for the diagnosis of lung cancer. J. Clin. Oncol. 2007, 25, 5578-5583.

- Zhang, Z., Bast, R. C. Jr., Yu, Y., Li, J. et al., Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res. 2004, 64, 5882-5890.
- Liao, H., Wu, J., Kuhn, E., Chin, W. et al, Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. Arthritis Rheum. 2004, 50, 3792-3803.
- Rithidech, K. N., Honikel, L., Milazzo, M., Madigan, D. et al., Protein expression profiles in pediatric multiple sclerosis: potential biomarkers. Mult. Scler. 2009, 15, 455-464.
- Meuwis, M. A., Fillet, M., Geurts, P., de Seny, D. et al., Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling. Biochem. Pharmacol. 2007, 73, 1422-1433.
- He, Q. Y., Lau, G. K. K., Zhou, Y., Yuen, S. T. et al., Serum biomarkers of hepatitis B virus infected liver inflammation: A proteomic study. Proteomics 2003, 3, 666–674.
- Thayan, R., Huat, T. L., Seed, L. L. C., Tan, C. P. L. et al., The use of two-dimension electrophoresis to identify serum biomarkers from patients with dengue haemorrhagic fever. Trans. R. Soc. Trop. Med. Hyg. 2009, 103, 413-419.
- Ren, Y., He, Q. Y., Fan, J., Jones, B. et al., The use of proteomics in the discovery of serum biomarkers from patients with severe acute respiratory syndrome. Proteomics. 2004, 4, 3477-3484.
- Kassa, F. A., Shio, M. T., Bellemare, M. J., Faye, B. et al., New inflammation-related biomarkers during malaria infection. PLoS One. 2011, 6, e26495.
- Weinberg, E. O., Shimpo, M., Hurwitz, S., Tominaga, S. et al., Identification of serum soluble ST2 receptor as a novel heart failure biomarker. Circulation. 2003, 107, 721-726.
- Ladd, J., Lu, H., Taylor, A. D., Goodell, V. et al., Direct detection of carcinoembryonic antigen autoantibodies in clinical human serum samples using a surface plasmon resonance sensor. Colloids Surf. B. Biointerfaces 2009, 70, 1-6.

- Chua, J. H., Chee, R. E., Agarwal, A., Wong, S. M., Zhang, G. J., Label-free electrical detection of cardiac biomarker with complementary metal-oxide semiconductor-compatible silicon nanowire sensor arrays. Anal Chem. 2009, 81, 6266-6271.