We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,400
Open access books available

118,000
International authors and editors

130M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

17

Serum Peptidomics

Kaihua Wei[1], Qingwei Ma[2], Yunbo Sun[1], Xiaoming Zhou[1], Weirong Guo[2] and Jian Yuan[1]
[1]*State Key Lab of Proteomics, Beijing Proteome Research Center,*
[2]*Bioyong (Beijing) Technologies Limited, Beijing, China*

1. Introduction

1.1 Concept

1.1.1 Proteomics

The performer of life functions is the dynamically-changing protein, rather than the relatively-static gene. Accordingly, the study of protein is of practical significance in the explication of vital phenomena, especially in the revelation of onset, deterioration and outcome of human diseases, which is also a driving force for the emergence of proteomics. Wilkins and Williams (Wasinger VC. et al, 1995) initiated the study of proteomics by putting forward the concept of proteomics for the first time. The proteome is the set of expressed proteins in a given type of cells, tissues or an organism at a given time under defined conditions. Proteomics is the large-scale study of proteomes, to discover composition and expression of proteins in organism, to understand interactions between proteins and explore functions of proteins and laws of vital activities of cells. It covers expression proteomics, functional proteomics and cell-localization proteomics etc. The proteomics technologies has provided a new tool for studying the biomarkers, pathogeny mechanism, diagnostic methods of diseases.

1.1.2 Serum proteomics

When proteomics comes to the clinical applications, it mainly refers to serum proteomics. The features of serum proteome research: firstly, it is easily to access samples, which means that it is able to meet the research requirements and easy to standardize; secondly, the dynamic variation in serum proteins is capable to reflect the pathological changing state of human organs; this is of far-reaching importance for disease diagnosis and curative effect monitoring. Human Proteome Organization (HUPO) brought the human plasma/serum proteome plan under the first-phase of the human proteome plans, which is showing the significance of studying serum proteomics for healthy and sick states of human.

Taking all proteins expressed in the serum of selected target clusters as the object, the serum proteomics, based on the normal protein expression profiles, aims to look for the differential proteins and define disease-associated proteins, the structures and functions of which will further be studied. In the hope of presenting a new approach for studying the pathological and physiological mechanisms of severe diseases, specific protein markers are expected to

be found for early diagnosis and drug targets. Compared with the tissue proteins and cell proteins, serum proteins are unique in many aspects such as the largest number (tens of thousands of varieties), extremely great difference in contents of proteins (with a difference of more than 10^8-10^{12}), extremely instable variety and content of low-abundance proteins. The main components of serum proteins are albumin and immunoglobulin, which share the features like high abundance, large molecular weight and easy-to-detect. Besides, non-protein substances like lipids and salts also exist in the serum. All other components will interfere with the study of small-molecular-weight proteins in the serum. Usually, all the important information come from the low-abundance proteins and small-molecular-weight proteins with great varieties and different properties, on the other hand, the features of these proteins make them difficult to be separated and identified.

1.1.3 Serum peptidomics

With the deepening of biological study of proteins, a kind of non-protein intermediate that composes of amino acid is discovered and described as polypeptide. According to biochemists, peptides are short polymers of amino acids linked by amino bonds (also called peptide bonds). Polypeptide is a kind of peptides with more than 10 amino acids (if fewer than 10, they are called oligo-peptides). In our study here, proteins with the molecular weight less than 10KDa fall into the category of the polypeptides. Besides molecular weights, the polypeptides are also different from proteins in functions: firstly, the polypeptide are information messenger, which can arouse various physiological activities and regulate biochemical response; secondly, polypeptide have high bioactivity; thirdly, as smaller molecule, the structure of polypeptide is easier to rebuild and simpler for artificial synthesis and chemosynthesis; fourthly, fragments of polypeptide can be used for further research of protein features and changing and synthesizing proteins as basic materials. The polypeptide is a sort of biologically active substance related to cell functions of organisms. Tens of thousands of polypeptides have been found existing in organisms and can be synthesized in all cells. Moreover, nearly all cells are regulated by polypeptides, which play a role in hormone, nerves, cell growth and reproduction, etc. Presently, in company with the instant development of proteomics and mass-spectrometric technology, more and more researchers are turning their eyes to polypeptide, which thereby bring about the proteomics-based peptidomics.

The main clinical application of peptidomics is the serum peptidomics. The serum peptidome, generally referring to serum peptidome profiling, is to detect the accurate mass value of polypeptide in serum by mass spectrometry and process mass spectrum with the bioinformatics method to build a polypeptide profiling. In the profiling, the peptidome identified by an accurate mass value can be further analyzed into amino acid sequence through tandem mass spectrometer, thereby used to identify the precursor proteins and their biogenetic derivation. Through contrasting differences in the serum peptidome profiling of the patients and the healthy controls, proteins or polypeptides specifically expressed in disease state and the biomarkers associated with diseases can be discovered to perform the studies of proteins related to early diagnosis, classification and subtype, onset mechanism of diseases. The serum polypeptide profiling is reputed to be a "brand-new health fingerprinting library" technology (Tarnaris A et al, 2006) and has become a research hotspot of proteomics.

1.2 History of serum peptidomics platform

Such are three major technical approaches for proteomics study as two-dimensional gel electrophoresis (2-DE), mass spectrometry (MS) and bioinformatics. Similar to the proteomics technology, serum peptidomics technology also involves in technical advancement and upgrade in sample preparation, detection and result analysis, which jointly constitute a platform in serum peptidomics research. Most of studies rely on matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS). The working principle of MALDI-TOF-MS is: the pulse laser is utilized to force the matrix to absorb energy from laser so that the solid-phase polypeptide samples are ionized; the ionized peptides are put into the TOF mass analyzer and separated due to difference in the mass-to-charge ratio (m/z); the peptide mass fingerprint (PMF), peptide sequence tag (PST) or the amino acid sequences can be obtained through measuring peptide ions; qualitative identification or quantitative analysis of peptides can be accomplished through searching in the proteome database with corresponding software. On this basis, three serum peptidomics research platforms have been developed successfully.

1.2.1 The 1st generation of serum peptidomics platform: SELDI

Different from MALDI-TOF-MS, the surface enhanced laser desorption/ionization(SELDI) time of flight mass spectrometry platform combines the protein chip with multiple high technologies (highly integrated, ultra-micro, computerized, automat-zed) based on the chromatogram principle, thereby to strengthen the affinity and capture ability of the chip so that proteins are selectively absorbed by the chemically-modified solid surface. The proteins captured by the chip surface are ionized in the ion source and their weight is detected by referring to the different flight time in the flight tube. In this platform, protein chip is the core of the whole system.

Distinguished by chemical modification, the protein chip is divided into hydrophobic surface (H4), normal phase (NP), weak cation exchange (WCX), strong anion exchange (SAX) and immobilized metal affinity capture (IMAC) which are aim to fit for different detecting requirement. Different chemisorption media allows an extremely large amount of proteins to be reduced to a relatively low level, those remained proteins can be absorbed by the surface of chip. They are characterized by: (1) can be directly used for crude biological samples, such as serum, urine, body fluid and cell lysis solution, etc. (2) low dosage of sample is required, generally is 0.5-15μl or 2000 cells; (3) high throughput with automatic operation; (4) rapid discovery for multiple biomarkers and some low-abundance, small-molecular-weight proteins; (5)high sensitivity lower the limitation of detection (LOD) to 1fmol (10^{-15}mol); (6) special identification ability for hydrophobic protein, especially for membrane proteins; (7) a high-efficiency and cost-effective system integrates all the protein separation, purification, identification, detection and data analysis process.

The processing and analysis system produces the results in forms like scanning profiling, bar graph and electrophoresis patterns (simulation gel image) and analyses the difference among two or more groups of results to find out the special mass spectra of identification information. The procedure includes: (1) database building; (2) internal and external information calibration; (3) data processing and analysis.

The data processing of serum protein profiles commonly includes peak detection, data grouping, marker selection, marker evaluation and composite mode building. In information analysis and processing, two algorithms will be used: (1) non-monitoring algorithm, like self-organizing cluster analysis; (2) monitoring algorithm, like artificial neural network (ANN). These two, especially the latter one, are more and more found in studies regarding tumor diagnosis with satisfying results.

SELDI, however, has some disadvantages: it is incapable of identifying differential proteins screened online; in the profiling, the peak height and protein concentration not always have a linear relationship; in the detection process, many factors are involved but the control method hasn't been standardized yet, which leads to poor reproduction quality of characteristic peaks of the same disease among different researchers.

1.2.2 The 2nd generation of serum peptidomics research technology system: ClinProt

To overcome the defects of SELDI, the ClinProt platform with functional magnetic beads, AnchorChip technology, MALDI-TOF-MS and ClinProTools has been developed. This platform has magnetic bead separating system, MS detecting system, analysis software and optional body fluid sample automatic processing system. The basic procedure is as follows: first of all, samples like serum, plasma, urine and cerebrospinal fluid, etc. from patients or the healthy controls have their high-abundance proteins and other impurities like salts to be eliminated and have the low-abundance proteins to be enriched with the magnetic bead. After separation and purification, samples are mixed with matrixes and directly drip on the AnchorChip target plate, and then protein mass spectrogram can be obtained through time-of-flight-MS (TOF-MS). The analysis software is employed to compare the differential expressed proteins of the patients or the healthy controls to get the specific mass spectra of the both for predicting the category of unknown samples (sick or healthy). Therefore, the core of this platform is the magnetic bead system and the software system.

Magnetic reagents include magnetic beads based Weak Cation Exchange (MB-WCX), Immobilized Metal Ion Affinity Chromatography Cu (MB-IMAC Cu) and Reverse Phase C18 (MB-RPC18), etc. The principle of magnetic beads is that the target proteins firstly combine with the magnetic microsphere surface ligand reversibility; secondly they will move in an outer magnetic field with a set direction by utilizing the magnetism of the magnetic beads to rapidly separate from surrounding medium. Compared with conventional technologies, the features of magnetic beads are: with large total surface area of magnetic beads, the specific basic groups on the surface can sufficiently combine with the low-abundance proteins in the serum to enhance the varieties of proteins, thereby to ensure the favorable specificity of the system; with simple and rapid operation, preliminary treatment can be finished through the simple blending, washing and elution process, which is suitable for clinical examination; the magnetic bead system has very high repetitiveness due to its large surface area; the liquid automatic processing system is a high throughput platform, it is capable to handle as many as 30,000 samples per day.

ClinProTools is a bioinformatics software boasting of all functions of biomarker detection and evaluation and capable of performing pre-processing of data, obtaining powerful, intuitive and visualized data from massive sample groups, carrying out peak statistics and pattern recognition, cluster analysis and independent test of data sensitivity, cross validation of specificity and meeting the classification demand of unknown samples.

Compared with SELDI, the magnetic bead particles have a larger total surface area due to their shapes and have stronger combining capacity, higher sensitivity and accuracy, so they perfectly cater for the research requirement of serum small-molecular proteins. Furthermore, biomarkers combined with the magnetic beads can be eluted, so this technology is suitable for MS and satisfy the requirement of sequential analysis in the further study. ClinProt system is able to perform the sequence identification task for the differentially expressed polypeptides/proteins to clearly distinct whether they are known or not.

1.2.3 The 3rd generation of serum peptidomics research platform: ClinTOF

Considering features of the serum polypeptide profiling and defects of the ClinProt system, it is fair to say that ClinTOF is the representative of the latest platform up to now.

On one hand, ClinTOF is capable of detecting the biomarker pattern or biomarker profiling indicating specific diseases in the biological liquid; on the other hand, this technology can identify the candidate for a single biomarker. ClinTOF composes of three parts, including magnetic beads, time-of-flight mass spectrometer (TOF MS) and analysis software BioExploer ™.

1. Magnetic Beads

The magnetic beads includes hydrophobic magnetic beads, metal affinity magnetic beads, ion exchange magnetic beads, glycoprotein magnetic beads and immunoaffinity magnetic beads. Further, SPE-C magnetic beads, the reagent dedicated for the serum polypeptide fingerprint diagnosis is available. Presently, magnetic beads have been found in biomedicine fields like immuno-magnetic separation (IMS), cell and cell organelle separation, microorganism detection and nucleic acid hybridization. The whole system is used in the clinical research of serum peptidomics.

2. TOF MS

Time-of-flight mass spectrometer (TOF MS) is used to obtain mass ratio and content of proteins captured by magnetic beads. ClinTOF, the clinical mass spectrometer, has adopted the cutting-edge 60Hz pulsed nitrogen laser which allows the data generated at the fastest speed compared with its counterparts. The zoom optics technology is employed with the laser speckle ranging from 50μm to 200 μm (adjustable), also the greatest adjustable range among like products, so that the size of laser speckles can meet the demands of different samples. The unique gap design in ion source has been utilized, keeping ion sources from contaminations and greatly reducing maintenance frequency. The laser system has been added with the laser energy leveling function, presenting more stable light intensity of the laser and more accurate data. The unique touch screen design has integrated MALDI-TOF control system and the PC system, making operations more simple and handy.

3. Analysis Software BioExploer™

BioExploer™ software is used both in processing genetic data and also protein data. BioExploer™ has combined visualized analysis and multiple mathematical algorithms to build pattern recognition models for MS data classification and forecast, and hunt the disease markers from data. It can perform data visualization, data reduction and data

mining over various types of MS data and build category prediction models. This software features: data of multiple forms can be analyzed with this software even if they have converted their formats; signal spectrogram can be visualized through the virtual gel graph and stack diagram; wavelet transform is used to deal with the mass spectrogram, including baseline elimination, spectrogram smoothing, peak selection and normalization; random statistical analysis can be made on one and more groups of spectrogram protein peaks; compatibility analysis is designed for the pairing data; the building and verification of the pattern recognition models involve in optimistic algorithms including genetic algorithm, radial basis neural network and the support vector machine and allow users to select the modeling space; output of data analysis reports and backup and storage at any moment.

The BioExploer™ control and analysis system adopts the display forms of scanning profiling (Fig.1) and electrophoresis patterns (simulation gel graph) (Fig.2). Statistical analysis graphs include the 3D sample distribution diagram (Fig.3), typical value-variogram (Fig.4) and 3D stack diagram (Fig.5). The user may switch the three graphs by clicking the three buttons at the lower left corner.

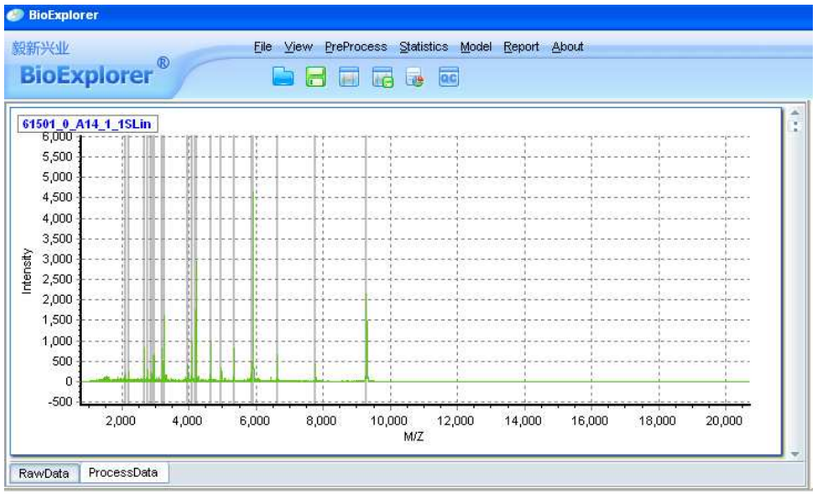


Fig. 1. ClinTOF System Scanning Profiling

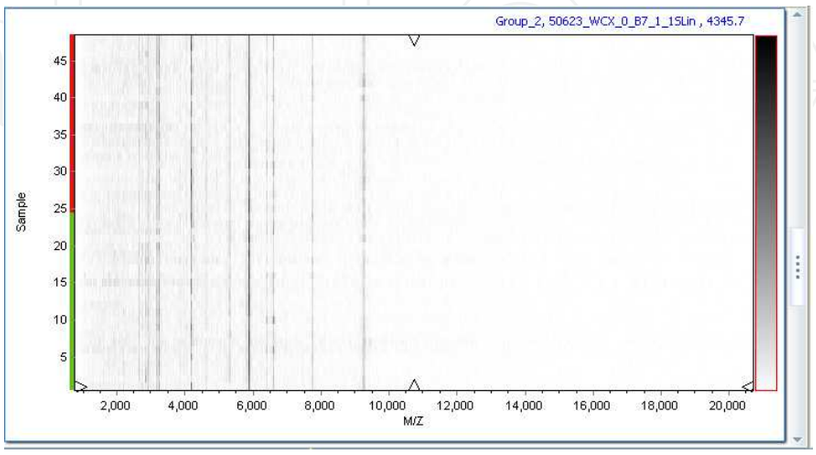


Fig. 2. Electrophoresis Graph of the ClinTOF System

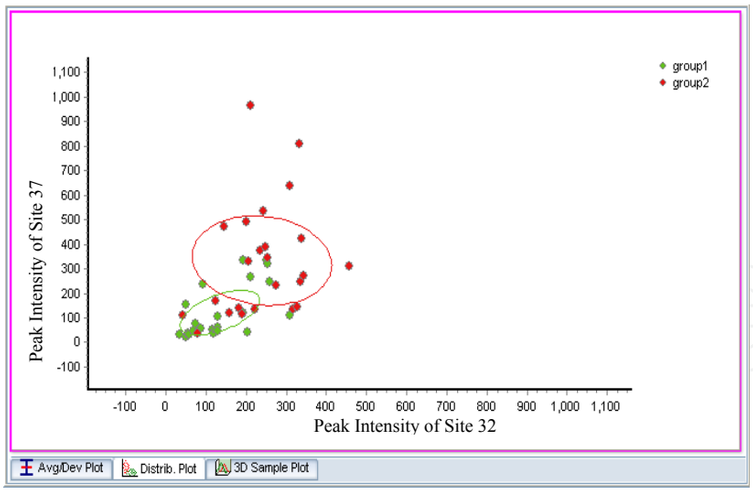


Fig. 3. ClinTOF System 3D Sample Distribution Diagram

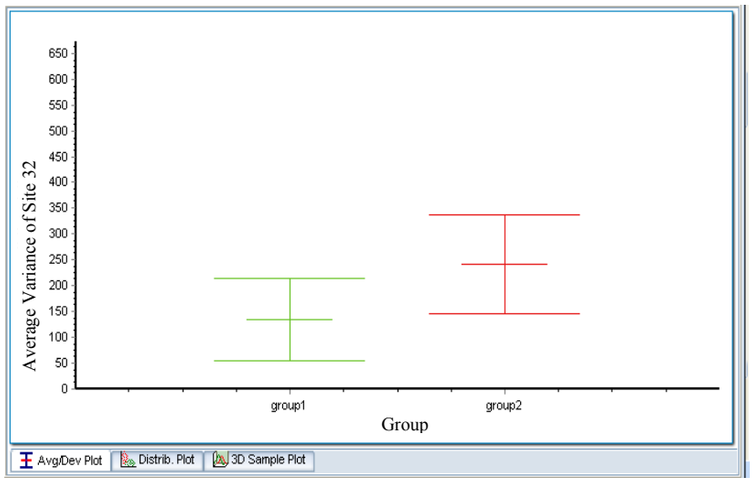


Fig. 4. ClinTOF System Typical Value Variogram

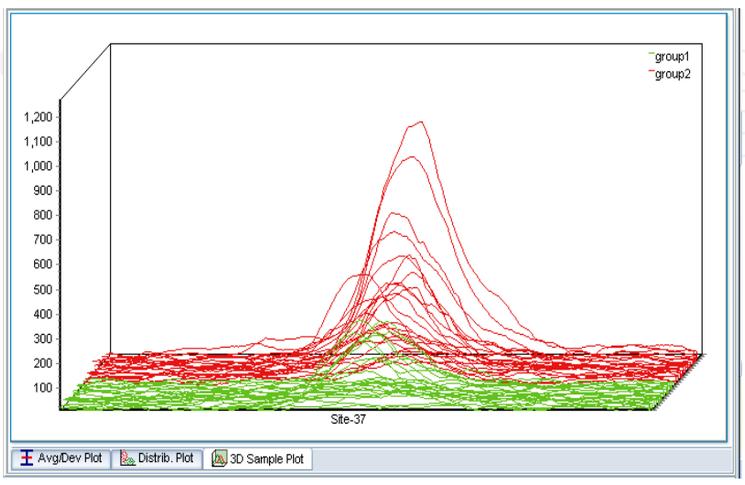


Fig. 5. ClinTOF System 3D Stack Diagram

4. Applications of ClinTOF Platform

ClinTOF can be used in disease peptidomics, early tumor diagnosis and curative effect evaluation, mental disease diagnosis, biomarker discovery, microorganism identification, single nucleotide polymorphism (SNP) detection and medicine quality control. This system as a breakthrough can detect more than 200 polypeptides at the same time by utilizing magnetic bead reagent, mass spectrometry and pattern analysis. BIOYONG is the only manufacturer in China for ClinTOF system development and certification and has obtained many patents. For the moment, this technology has been widely used in the clinical based researches, granted a solid scientific foundation for future promotion.

Besides the intrinsically strength of MALDI-TOF, the ClinTOF system has improved its stability and repeatability for clinical analysis. So far, it has detected more than 10,000 clinical samples and established detection models for colon and rectal cancers, lung cancer, hepatic carcinoma and brain glioma. Its accuracy for the early detection of cancers is above 85% and its specificity and sensitivity exceed 80%. MS models for some tumors based on the Clin TOF system have been built and many patents have gained authorization.

The ClinTOF system has been widely applied in studies on the early diagnosis of ovarian cancer, prostatic cancer, breast cancer, brain glioma, head and neck squamous cell carcinomas (HNSCC) and carcinoma of urinary bladder and disease diagnosis models have been built. For the colon and rectal cancers, the model is shown in Table 1. In the model built in the experiment, 70 of the normal cases and 60 of colon and rectal cancer cases are used; 31 of the normal cases and 35 of colon and rectal cancer cases are used; the sensitivity of the model is 84.17% and the specificity, 95.95%. For the model verified by blind samples, the normal/colon and rectal cancer cases is (31/35) and the sensitivity and specificity of the model both exceed 80%, with the accuracy reaching 84.85%. For the hepatic carcinoma, the diagnosis model is shown in Table 2. According to the established standard operating

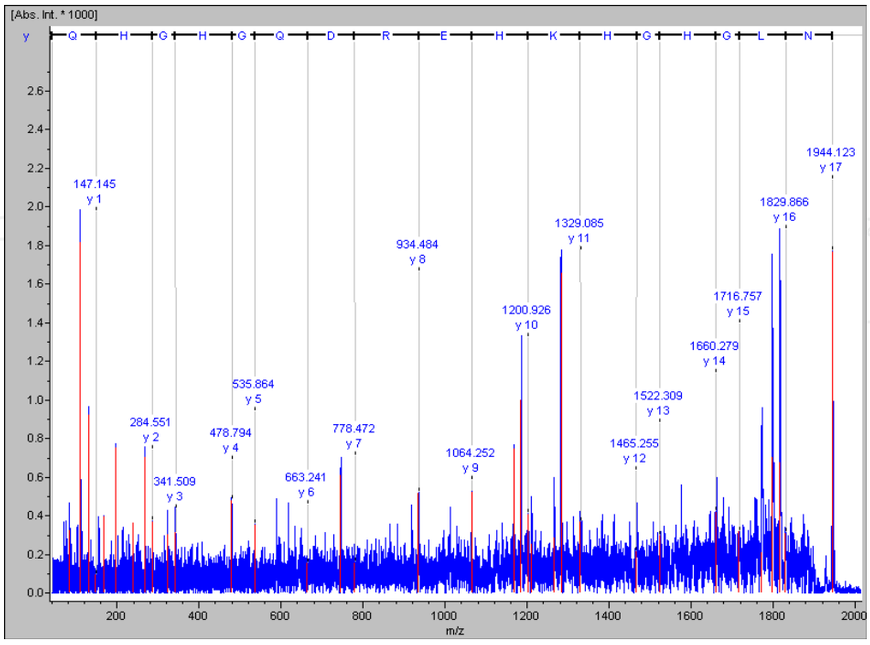


Fig. 6. Hepato Carcinoma Diagnosis Model of the ClinTOF System

procedure (SOP), BIOYONG has studied the polypeptide profiling of the 200 hepatic carcinoma cases and 200 normal serum cases. 28 differential polypeptides (P<0.0001) were obtained, with the molecular weight of 900 ~5000Da (Fig.6). Among them, 9 were down regulated (2.5~8 times) and 19 up-regulated (2.5~20 times). For the model built with SNN, the recognition rate is 100% and predictive ability, 98.39%. 100 samples were used for double-blind determination, with the correctness rate exceeding 90%.

	Training Set	Testing Set
Sensitivity	84.17%	77.4%(24/31)
Specificity	95.95%	91.4%(32/35)
Positive Predictive Value (PPV)	---	88.89%(24/27)
Negative Predictive Value (NPV)	---	82.1%(32/39)
Accuracy	---	84.85(56/66)

Table 1. Colon and Rectal Cancer Diagnosis Model of the ClinTOF System

Performance	Bruker	Bioyong	CipherGen	CipherGen
	microflex	ClinTOF	PBSII/C	PCS4000
Laser light source	nitrogen 337nm 1-20Hz adjustable	nitrogen 337nm 1-60Hz adjustable	nitrogen 337nm 10Hz	
Ion source technology	Delayed extraction technology			
Ion source pattern	Positive and negative ion sources			
Vacuum system capacity	10^{-6}pa		10^{-7}pa	10^{-4}pa
Detection system	Micro-channel plate detector		Electron multiplier	
	Shielding the "gate" function of noise peaks			
Mass range	>600KDa	>500KDa	>500KDa	>380KDa
Sensitivity	<1fmol		<10fmol	
Resolution	>3500FWHM	>2500FWHM	>700FWHM	>1000FWHM
Accuracy (inner calibration)	≤50ppm	≤50ppm	≤100ppm	
Accuracy (outer calibration)	≤150ppm	≤100ppm	≤2500ppm	≤250ppm
Other features	Patented AnchorChip MALDI sample target	Zoom optics technology, with the laser speckle ranging 50um to 200um adjustable	Coaxial laser technology	
	WishperMode technology reduces lab sound pollution	Unique remote ion source design of big gap, keeping ion source from contamination		
Applications	Proteomics, SNP, biomarker analysis	Proteomics, SNP, biomarker analysis, tissue imaging, microorganism, organ-small molecules, clinical examination in hospitals		

Table 2. Serum Peptidomics Research Technical Systems

1.2.4 Comparison of serum peptidomics research platform

The three systems, namely, SELDI system, ClinProt system and ClinTOF system, of the serum peptidomics research platform are compared in Table 2.

1.3 Main application areas of serum peptidomics

For the moment, the serum peptidomics is clinically used in complex diseases involved in polygene and featuring multi-cause heterogeneity, like cancer (including colon and rectal cancer, lung cancer, hepatic carcinoma, esophagus cancer, stomach cancer, cervical carcinoma and nasopharyngeal carcinoma), nerve degenerative diseases (including Alzheimer disease, Parkinson's disease, Huntington's Disease), autoimmune diseases (rheumatoid arthritis, system lupus erythematosus syndromes), cardio-cerebrovascular disease and palsy, to discover protein/polypeptide profiling and biomarker profiling as well as the single biomarker in the serum of sufferers of these diseases, making breakthroughs and providing a new tool for the researches of disease pathogeny, drug target, diagnosis and treatment.

2. Research method of serum peptidomics

The research of serum peptidomics involves in sample detection and data analysis. Fig.7 is the work flow of serum peptidomics. As shown in Fig.7, blood samples from the

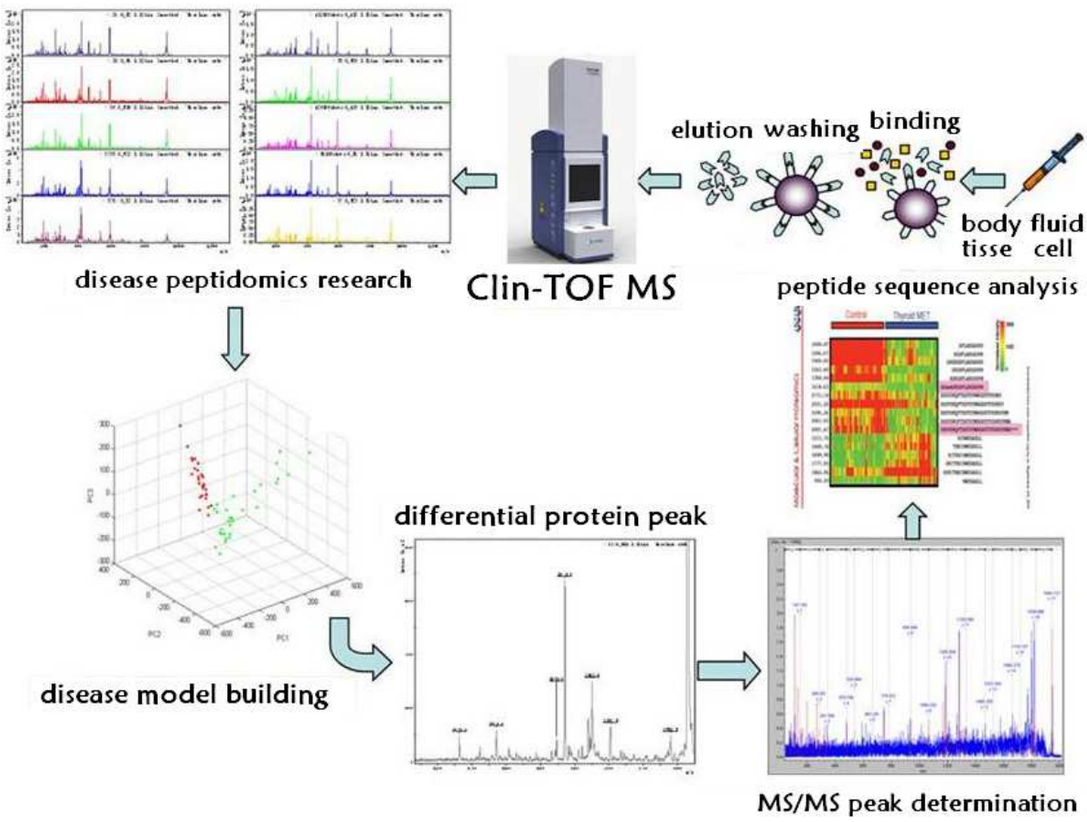


Fig. 7. Technical Process for Serum Peptidomics

pathological group and the healthy control group are collected and then the serum is separated from the blood. The serum is mixed with magnetic beads to extract serum polypeptides for detection with the mass spectrometry. The spectra obtained can produce characteristic spectrum peaks. The meaningful characteristic spectrum peaks are screened out with the statistical method. A prediction model is built with the pattern recognition method and validated with test data. After continuous optimization, a disease diagnosis model and a group of characteristic spectrum peaks can be obtained.

Briefly, the research includes magnetic beads, detection instrument, analysis software, polypeptide identification and clinical model.

2.1 Polypeptide extraction

The separation and purification methods adopted in the serum peptidomics research depend on the nature of the extracted substance. The commonly used methods for polypeptide extraction and separation include: salting removal method, ultra-filtration process, gel filtration, isoelectric point precipitation method, ion-exchange chromatography, affinity chromatography, adsorption chromatography, countercurrent distribution and enzymolysis method. These methods often work together to separate and purify specific substances.

2.1.1 High Pressure Liquid Chromatography (HPLC)

HPLC is a favorable method for peptides separation, because the HPLC can complete the separation in a short time under suitable chromatographic conditions, and more importantly, HPLC is capable of producing polypeptide of bioactivity at the preparative scale. Many scholars therefore have done substantive work in looking for the best conditions for separating and preparing polypeptide substances. How to maintain the activity of polypeptide, how to select stationary phase material and eluent type, how to make analytic determination are all contents of the present study. Common methods include: reversed phase high pressure liquid chromatography (RP-HPLC), hydrophobic interaction chromatography (HIC), Size-Exclusion chromatography (SEC), Ion-Exchange chromatography (IEC), Chromatography of Membrane Protein (CMP), High-Performance Displacement Chromatography (HPDC) and Perfusion Chromatography (PC).

2.1.2 Affinity Chromatography

Affinity Chromatography (AC) is the method of separating substances based on the specific affinity between ligand connecting to the stationary phase matrix and the ligand having interaction with the specificity. Since 1968 when Cuatrecasas put forward the concept of affinity chromatography, in searching for the specific affinity interaction substances many combinations have been found, like antigen-antibody, enzyme-substrate, agglutinin-polyose, oligonucleotides and their complementary strands. For the separation of polypeptide substances, currently the monoclonal antibody or biological simulation ligand can be used for affinity to such substances. These ligands can be natural or artificially synthesized according to their structure. Immobilized Metal Affinity Chromatography (IMAC) is an affinity method developed in recent years. Some metal ions are chelated on the stationary phase substrate, like Cu^{2+}, Ni^{2+} and Fe^{3+}. The magnetic beads can be chelated through the coordination bond to connect polypeptides that contain Lys, Met, Asp, Arg, Tyr, Glu and His on the side chain. In

particular, structures with the peptide sequences containing His-X-X-X-His are easiest combined to the metal ion affinity column, featuring good purification effect.

2.1.3 Capillary Electrophoresis

Capillary electrophoresis (CE) was invented by Hjerten at the 1960s on the basis of the conventional electrophoretic technology and improves the electrophoresis efficiency by dozens of times by replacing the large electrophoresis tanks with the capillaries. This technology developed rapidly from 1980 and became a good tool for separation and determining the nature of polypeptides and protein substances by bio-chemical analysts and biochemical scholars. By application principles, CE can be divided into: Capillary Zone Electrophoresis (CZE), Capillary Isoelectric Focusing (CIEF), Capillary Gel Electrophoresis (CGE) and Micellar Electrokinetic Electrophoresis Chromatography(MECC).

2.1.4 Solid phase protein chip technology

The solid phase protein chip technology usually refers to the sample separation part of the SELDI system. The SELDI can not only use the enzyme, antibody, receptor and DNA as the sorting basis of chips, but also take chemical mediators of different natures as the protein sorting basis according to the different chemical natures of proteins, thereby to reduce the requirement on the sample purification and enlarge the sorting range of samples. SELDI chips can be divided into the chemical surface chips and biological surface chips according to the detection purpose. The former include hydrophobic surface (H4), normal phase (NP), weak cation exchange (WCX), strong anion exchange (SAX) and immobilized metal affinity capture (IMAC) and the latter include antibody-antigen surface chips, receptor-ligand surface chips, enzyme-substrate surface chips and DNA-protein surface chips, specifically for detecting the corresponding polypeptide molecules.

2.1.5 Liquid phase protein chip technology – Magnetic bead

The magnetic bead is a new multifunctional reagent developed in recent years and widely used in biomedicine. Owing to diversified surfaces of macromolecular coat, magnetic beads can be coupled with various biologically active substances (like antibody, antigen, receptor, enzyme and nucleic acid), which can be fixed to the magnetic beads to further identify corresponding antigen or antibody, ligand, substrate, nucleic acid in the reaction medium, thereby to realize separation or detection. The results show magnetic beads have functions of both the carrier and separation performer. They can simplify complicated operations by utilizing the physical, chemical and biomedicine principles, thereby to greatly shorten the period of conventional test. Different from protein chips, magnetic beads are composed of many magnetic spherical granules, which provide them with greater surface area and enable them to combine more specific proteins and to have higher sensitivity and accuracy.

To be detailed, the disease associated differential protein liquid chip SPE-C magnetic beads; hydrophobic MB-HI C 1 (suitable for purification and enrichment of proteins more-than-20kDa), MB-HI C 3 (suitable for purification and enrichment of 8-20kDa proteins), MB-HI C 8 (suitable for purification and enrichment of 1-10 kDa proteins/polypeptide), MB-HI C 18 (suitable for purification and enrichment of less-than-4kDa polypeptide); metal affinity magnetic beads include MB-IMAC-Fe (for capture and enrichment of phosphorylation proteins) and MB-IMAC-Cu (for the capture and enrichment of specific affinity

proteins/polypeptides); ion exchange magnetic beads include MB-WCX (purification and enrichment of acid proteins with the cation-exchange chromatography technology) and MB-WAX (purification and enrichment of basic proteins with the anion-exchange chromatography technology); glycoprotein magnetic beads include ConA and ConB (for purification and enrichment of glycoprotein); the immune fishing liquid chips include immune affinity magnetic bead ProteinG (the ProteinG on the magnetic bead can combine with any one antibody, for screening specific antigens).

2.1.6 Systematic applications of polypeptide separation engineering

Polypeptide separation technologies mentioned above are combined to use in practice. Different separation tools will be used according to the nature of polypeptide. In particular in the post genome era, researches on the proteome are deepened and people are making continuous progress in tools for separating polypeptides and proteins. They have comprehensively utilized natures of proteins and polypeptides and adopted both the routine protein and polypeptide extraction methods aforementioned and also the high efficiency liquid chromatography, capillary electrophoresis and 2-d electrophoresis, to get as many polypeptides as possible.

2.2 Sample detection

2.2.1 Detection of N-terminal sequence with Edman degradation method

The Edman degradation method used for sequencing can obtain a precise peptide sequence, therefore making it a major method for protein identification. Yet its sequencing speed is slow and expansive. With technical breakthroughs in microsequencing and speed, Edman degradation method will exert a major role in proteome researches. The C-terminal Maxam-Gilbert method, similar to Edman, has been studied for years and produced automated analyzers, but its reaction efficiency is low and usually requires more samples.

2.2.2 Amino acid composition analysis

Amino acid composition analysis is frequently used for protein identification owing to its low cost. Different from the peptide mass or sequence tags, amino acid composition analysis identifies proteins by utilizing the specific amino acid component of different proteins. This method can be used for identifying 2-DE separated proteins. The radio-labeled amino acid is used to determine amino acid components of proteins, or the proteins are converted to the PVDF membrane and after the automatic derivation of amino acid, undergo chromatographic separation to obtain data. Then, inquiries are made to the database to rank proteins in the database by the amount of difference of two components, as a result, the top ranking proteins having greater reliability. Yet this method has some defects: slow speed and require a great amount of proteins or peptides; restricted in the ultramicro analysis;the possible amino acid variation due to the incomplete acidic hydrolysis or partial degradation.

2.2.3 Mass Spectrometry

Mass Spectrometry (MS): MS is an analytical technique that measures the mass-to-charge ratio (or mass) of charged particles, molecules or molecular fragments. MS provides information of molecular weight, molecular formula, isotopic element composition of

molecules and molecular structure of samples analyzed. It has been widely applied in protein and polypeptide analysis. In particular, it is suitable for the analysis and identification of polypeptide substances in the online analysis after separation and purification due to its hypersensitivity and rapidity. Commonly, MS includes electrospray MS (in the spray process, the continuous ionization method makes the polypeptide samples ionized), fast atom bombardment MS (FAB MS) and isotopic element MS. Among them, the Continuous-Flow Fast Atom Bombardment, cf-FAB) and the Electrospray Ionization (ESI) have just been developed in recent years.

Continuous-Flow Fast Atom Bombardment(cf-FAB): a kind of weak ionization technology, it is capable of ionizing peptides or small-molecular-weight proteins into the form of MH+ or (M-H). It is mostly applied in the separation and detection of peptides and has moderate resolution, with the accuracy greater than +0.2amu and flow rate of 0.5-1.5μl·Ml-1. In the determination, the mobile phase shall be added with 0.5%-10% substrate like glycerol and high organic solvents, so that samples can be sensitized at the detection probe. The cf-FAB is usually used together with HPLC and CEZ, to realize the purpose of isolation analysis. The cf-FAB analysis methods have been built for many polypeptides and well applied.

Electrospray Ionization (ESI): able to generate multivalent ionized proteins or polypeptides, allowing analyzing of proteins with the molecular weight reaching 100kD; its resolution is 1500-2000 amu and accuracy about 0.01 %. ESI is more suitable for the online analysis of proteins with large molecular weight and requires gasification or organic solvents for the sample sensitization. It has been a success to combine ESI and HPLC for separation and analysis of GH and hemoglobin. ESI can also be used together with CEZ.

MALDI-TOF MS: in this method, the ionization of polypeptide samples is realized with the substrate absorbing the laser energy. It is a tool for accurately determining the molecular mass in the current protein identification and particularly suitable for the determination of the molecular weight of mixed proteins and polypeptides, featuring high sensitivity and resolution. For the moment, it is a necessary tool for the proteomics research. Working with the coupling technique of the liquid chromatography, this method can identify polypeptides at a high efficiency. Especially, when MS technologies of different principles are coupled, they can not only obtain the molecular weight of polypeptides, but also determine the sequential structure. This technology will exert a decisive effect in the future proteomics study.

2.2.4 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR): NMR profiling has purely digital signals, excessive overlapping range (due to the large molecular weight) and weak signals, so it is seldom used in analysis of proteins and polypeptides. In company with the application of 2D, 3D and 4D NMR and the progress of molecular biology and computer processing technology, NMR has gradually become a main approach for analysis of proteins and polypeptides. NMR can be used for determining amino acid sequence and the content of components in mixtures. Yet some problems need solving if this method is used for protein analysis, for instance, how to give proteins with large molecular weight a specific shape to facilitate quantitative and qualitative analysis and how to reduce the data processing time, which are being studied by many scholars. Despite of its infrequent use in the protein analysis, NMR is extremely useful in analyzing small peptides with the molecules having less than 30 amino acids, in which case, it can overcome the foregoing defects and realize rapid and accurate analysis.

2.2.5 Others

Besides the foregoing methods, amino acid composition analysis, amino acid sequence analysis, Field Desorption Mass Spectrometer (FDMS), IR, UV spectra, circular dichroism spectrum, bioassay technique, tagging method and immunologic method have also been used for the result identification, analysis and detection of polypeptides.

2.3 Analysis method

Regarding research on serum peptidomics, the analysis of data collected shall have: high-efficient analysis technical platform, computer and network have become a necessary tool of biological research; high throughput technical platform, mainly targeting at how to use the information technology to analyze the giant data; data mining technical platform, which shall be able to mine knowledge from the massive data saved in database or other information banks for the analysis; data visualization technical platform, in describing the systematic relations, the functions of nucleic acid, protein, cell, organ and tissue shall be considered, that is to say, a systematic method shall be used to learn about vital activities.

Currently, databases used for proteomics research include SWISS-PROT, BLOCKS, SMART, PROSITE, WORLD-2D-PAGE, EMBL, GenBank, DDBJ, ProClass, PR INTS, MASCOT, PROTO-MAP, DOMO, PDB and NCBI. Among them, SWISS-PROT is a real protein sequence database and also the largest and most diversified proteome database in the world. EMBL is to collect protein sequences that have been translated from nucleic acid automatically and not yet entered the SWISS-PROT. NCBI contains protein sequences translated from DNA in the GenBank and from the PDB, SWISS-PROT and PIR databases.

Presently, many tools and methods used for MS data process and analysis of serum peptidomics have been developed, mainly including:

2.3.1 Bioconductor

Bioconductor is an open source and open development software project, with the broad goals of providing widespread access to a broad range of powerful statistical and graphical methods for the analysis of genomic data, facilitating the inclusion of biological metadata and driving comprehensive analysis and application of data. Its application function is to provide users with the integrated packages. It has also provided many packages of MS data processing and analysis for users. Bioconductor is based on the R language, so it requires that users must be familiar with the R-language working environment and have some knowledge of programming, that is to say, it will be difficult for the general clinical and lab workers.

2.3.2 MATLAB

MATLAB is a piece of commercial software integrating statistical analysis and engineering computation. Taking MATLAB as the development platform, it will be possible to realize the pretreatment, display and statistical analysis of MS data. In the research of serum peptidome profiling of prostatic cancer, breast cancer and bladder cancer, this tool together with the GENESPRING of Agilent has achieved good results in data analysis (Villanueva et al, 2005).

2.3.3 TOF-MS Based Software

1. Spectrogram Pretreatment

Due to many influencing factors in the MS experiment, the original spectrogram produced by MS must be pretreated to eliminate disturbance. Pretreatment of the MS data includes baseline elimination, filtration and noise elimination, standardization, peak detection and peak quantification. The comparison of commonly-used pretreatment methods and tools are shown in Table 3 (Cruz-Marcelo et al, 2008).

Algorithm and Tool	Main Functions	Relevant Information
ProteinChip Software 3.1 and Biomarker Wizard	Commercial software of Ciphergen Biosystems, designed for analyzing SELDI-TOF MS data	http://www.vermillion.com/
PROcess	An R-language based package of BioConductor, designed for pretreatment of SELDI-TOF MS	http://www.bioconductor.org/packages/bioc/1.8/html/PROcess.html
Cromwell	MatLab script to realize MS data pretreatment	http://bioinformatics.mdanderson.org/cromwell.html
SpecAlign	MS data pretreatment and peak alignment	http://physchem.ox.ac.uk/~jwong/specalign/index.htm
MassSpecWavelet	An R-language based package of BioConductor, using continuous wavelet transform for peak detection	http://www.bioconductor.org/packages/2.0/bioc/html/MassSpecWavelet.html

Table 3. Commonly-used MS Pretreatment Algorithms and Tools

2. Peak Alignment

The spectrograms after pretreatment shall undergo peak alignment. Many spectrograms are combined into a matrix file similar to the gene expression profile, namely the serum peptidome profiling. In the profiling, the line represents the peak of some specific charge-mass ratio (m/z), namely, the relative content of specific proteins or polypeptides, and the column represents samples. This is the foundation for follow-up bioinformatics analysis and its data quality directly influences the analysis results.

3. Bioinformatics Analysis

Usually, the first step is to make cluster analysis, mainly including shortest distance method, longest distance method, median method, centroid method, average linkage and Ward's minimum-variance method. Also, the data classification is carried out. The commonly-used methods include support vector machine (SVM), decision tree, neural networks and k nearest neighbor (kNN).

The research process based on serum peptidomics classification is first to divide the mass spectrometric data after pretreatment into a group of modeling data and the other group of validation data. The modeling data fall into the training set and testing set. Then, analysis

will be performed over the training set with t-test, Pearson correlation analysis and genetic algorithm, to find peaks of higher specificity to build a sorter. Then, the testing set makes tests, which shall be repeated and optimized. Finally, the validating data are used for validating to get a stable model.

4. TOF-MS System Based Softwares

Softwares based on the TOF-MS system mainly includes: (1) for the SELDI system, the ProteinChip Software 3.1 and Biomarker Wizard taking the decision tree as the core; (2) for the ClinProt system, the ClinProTools software taking cluster analysis as the core; (3) for the ClinTOF system, the BioExploer™ software taking specific vector machine (SVM), decision, tree, neural networks and k nearest neighbor (kNN) as the core.

Other bioinformatics tools for MS correlation analysis include: MapQuant, MASPECTRAS, SpecArray, msInspect and MZMine. These tools or softwares haven't realized seamless connection with serum peptidomics data, so they fail to perfectly accomplish the data management and analysis based on MS serum peptidome profiling.

3. Disease serum peptidomics research progress

Nowadays in clinical detection, the serum biochemical indicators fail to accomplish the task of diagnosing complex diseases. For instance, there are only one or two serological diagnosis indicators available for specific cancers; even worse, these indicators cannot diagnose diseases independently and sometimes confuse the cancer with benign tumors or inflammation. For example, the prostate specific antigen (PSA) is a major diagnosis indicator for prostatic cancer, but not a biomarker for the prostatic cancer specificity, for the reason that 15%~25% prostatic cancer patients see their serum PSA falling into the normal range. Besides, PSA may also rise due to the benign hyperplasia of prostate, urinary infection, acute prostatitis, retention of urine and the per rectum operations. Accordingly, serum peptidomics analysis, as a new clinical diagnosis method, has made great progress in early diagnosis of complex diseases like tumor, neurological degenerative diseases and autoimmune disease.

3.1 Tumor

3.1.1 Ovarian cancer

Ovarian cancer is the most maligne tumor among all malignant tumors of the reproductive system, with the pathogenic factors unclear yet but possibly relating to the reproductive and hereditary factors. Most ovarian cancer cases are detected in the late stage and seldom cured. For the moment, the universally-ratified specificity biomarker CA125 for ovarian cancer diagnosis, as the single biomarker, sees the positive predicted value less than 10%. Therefore it is urgent to develop an approach for the early-stage clinical diagnosis so as to enhance the survival rate. The serum based peptidomics has seen great progress in researches of ovarian cancer diagnosis.

Pleasantly, based on findings of the SELDI system, the US Vermilion Company and Quest Diagnostics worked together to develop the OVA1, which can determine the onset of ovarian cancer by detecting the cavum pelvis enclosed mass and decide whether operations

are needed, what operations are needed and who shall perform the operations if applicable. OVA1 ovarian cancer qualitative serum test forms a single digital grading system by combining the five immunoassay combinations. The test has shown that female older than 18, requiring operations of ovarian enclosed mass and not examined by the oncologist can undergo examination with OVA1. This immune test has determined five tried and true biomarkers, namely, thyroxin, apolipoprotein A-1, β2-microglobulin, transferin and cancer antigen 125. One algorithm is decided to determine the probability of tumor onset of female cavum pelvis enclosed mass. Quest Diagnostics exclusively provides OVA1 to the US clinical lab for three years. Food and Drug Administration (FDA) has approved OVA1 to be used for the high-sensitivity testing of ovarian cancer, with effect better than biopsy or operation examination, even if the radiation test results cannot show whether malignant tumors exist. Vermilion is devoted to discovering, developing and commercializing new-style high-value diagnostic tests, to help doctors diagnose, cure and perfect prognoses of the patient. Vermilion provides approaches for diagnosis of tumor, blood and heart disease and for ensuring women health (OVA1, Fremont, CA: Vermillion, Inc; 2009).

Petricoin et al., with the fund from the clinical proteome plan of the US FDA/NIH, has successfully applied the serum polypeptide profiling technology to make diagnosis of early-stage ovarian cancer. The ovarian cancer serum proteomics findings with the hydrophobic chips were reported: they discover that the content of five types of proteins in the serum of ovarian cancer patients and the healthy people is changing at the same time, which is of far reaching importance to the diagnosis of ovarian cancer. This proteome worked as the specificity biomarker for diagnosing ovarian cancer and double-blind researches were done to the serum of the healthy people, ovarian cancer patients and ovarian benign pathological changes, with results showing that the sensitivity of this method is 100%, specificity 95% and positive predicted value 94%; by contrast, the positive predicted value of CA125 was only 35% (Petricoin et al, 2002).

Katherine et al. utilized the protein fingerprint technology to load 184 serum samples on the strong anion exchange (SAX) chips. The 184 cases included 109 ovarian cancer cases, 19 benign ovarian tumor cases and 56 healthy people. The univariate and multivariate statistics method has been applied for analysis. From the protein fragments from the 140 serum samples, 3 groups of protein markers of diagnosis significance were obtained: the first group included 5 candidate protein markers, with the ovarian cancer sensitivity of 95.7%, specificity of 82.6% and accuracy of 89.2%; the other two groups of protein markers included 5 and 4 candidate protein markers respectively, with the sensitivity of 81.7% and 72.8%, specificity 94.9% and of 94.9% and accuracy of 88.2% and 83.9% respectively. After screening out 3 groups of protein markers, the rest 44 unknown serum samples were used for blind validation, showing that when the three groups were employed together, 41 serum cases were diagnosed correctly, to be detailed, 21 ovarian cancers were correctly diagnosed, 11 ovarian cancers of the progressive stage were correctly diagnosed and 10 of the 11 ovarian cancers of the early stage were confirmed; 6 low-malignancy potential tumors were diagnosed, 5 of the 6 benign tumors were excluded the possibility of ovarian cancer, 1 of the 10 serum cases of healthy people had a wrong diagnosis. In this sense, Katherine et al. believed that they had found ovarian cancer protein marker groups of diagnosis significance and can effectively distinguish healthy people from the benign/malignant ovarian tumors.

Ye et al. utilized the protein chip technology to analyze the serum of 80 ovarian cancer patients and 91 healthy people, discovering the differential protein peak at the position with

the molecular mass of 11700 Da, which was obviously higher than the density peak value of the control group. Then, the affinity chromatography method was used for purification of it and the protein sequence was then determined with the fluid chromatography and mass spectrometer. A polypeptide chain was synthesized. Finally, it was identified that this polypeptide fragment was haptoglobin chain and corresponding antibody was derived. Combined with CA125, this method saw the ovarian tumor diagnosis sensitivity of 95.7%, specificity 82.6%, enhanced the early diagnosis rate of ovarian cancers and lowered false positive of CA125. Currently, relevant technologies started being applied in screening ovarian cancers.

In addition, theses published on Lancet indicated the MS technology was used to analyze the composition pattern of serum peptidome and based on the pattern, found out differential points for detection of ovarian cancer clinically. Results showed that 50 ovarian cancer patients were all detected, with 18 patients suffering stage-I ovarian cancer; 63 of 66 gynecologic benign tumors were diagnosed correctly. In detection of ovarian cancer, the sensitivity, specificity and positive predicted value of this method were 100%, 95% and 94% respectively, much better than the conventional CA125 detection. In particular, it was successful in diagnosing the early-stage ovarian cancer (stage I), indicating that this technology may be used for the early-stage or early warning detection of ovarian cancer hopefully. Thereafter, a series of theses on the serum peptide profiling technology for disease diagnosis were published in the world. For instance, in 2004, Soltys et al. with the Stanford University published results of using serum peptidome profiling to diagnose squamous carcinoma on Clinical Cancer Research; in 2005, Kaz'ufumiHonda et al. with the Japanese National Cancer Center published thesis on utilizing low-resolution MS and high-resolution MS to diagnose pancreatic cancer on Cancer Research. At the beginning of 2006, Villanueva et al. with the US Sloan-Kettering Cancer Center published an article on using serum peptide profiling technology to diagnose bladder cancer, breast cancer and prostatic cancer; at the end of 2006, Nature published remarks on this article, signaling its support for this technical development and application. Lately, Lancet published an article on using the serum peptide profiling technology to diagnose tuberculosis, which produced sensitivity and specificity greater than 950k.

3.1.2 Breast cancer

We employed the ClinTOF system in the research of breast cancer, with the experimental design as shown in Fig.8. The serum polypeptides were extracted with magnetic beads and then the polypeptide profiling was obtained with MS detection, to build a disease model. Significant difference was found with the software analysis and the reliability of the model was validated with a certain number of blind samples. At the same time, the reliability of the difference was confirmed. Multiple tandem mass spectrometry were used to identify differential polypeptides. FlexAnalysss2.4, BioExplorer, SIMCAP+ and ClinProTools2.0 were used to analyze results, as shown in Fig.9, finding out 4 differential peaks; identification and functional forecast were performed on differential polypeptides.

Most of the polypeptide sequences identified lacked of only one or several amino acids. After comparing them with results and discussions in literatures, it was concluded that under different disease response conditions, the activity of external peptidase would undergo changes, thereby to obtain a series of polypeptides by cutting different protein loca. These polypeptides might share the same modif.

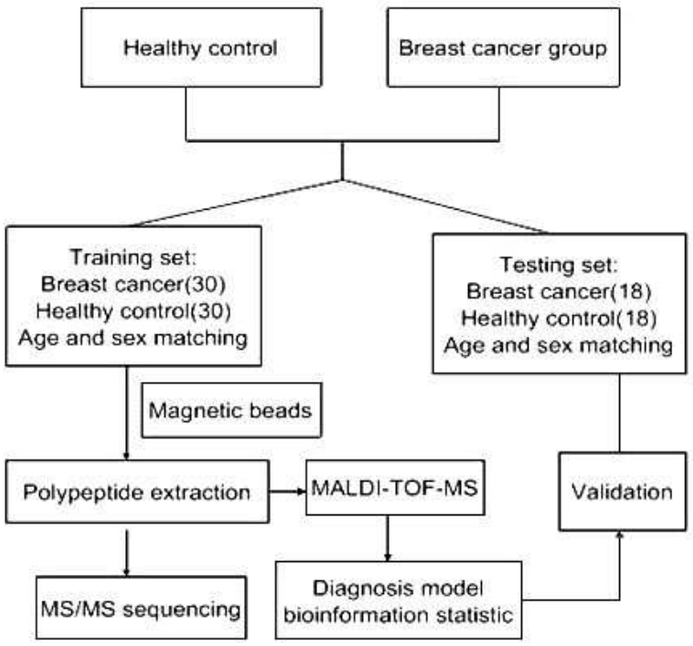
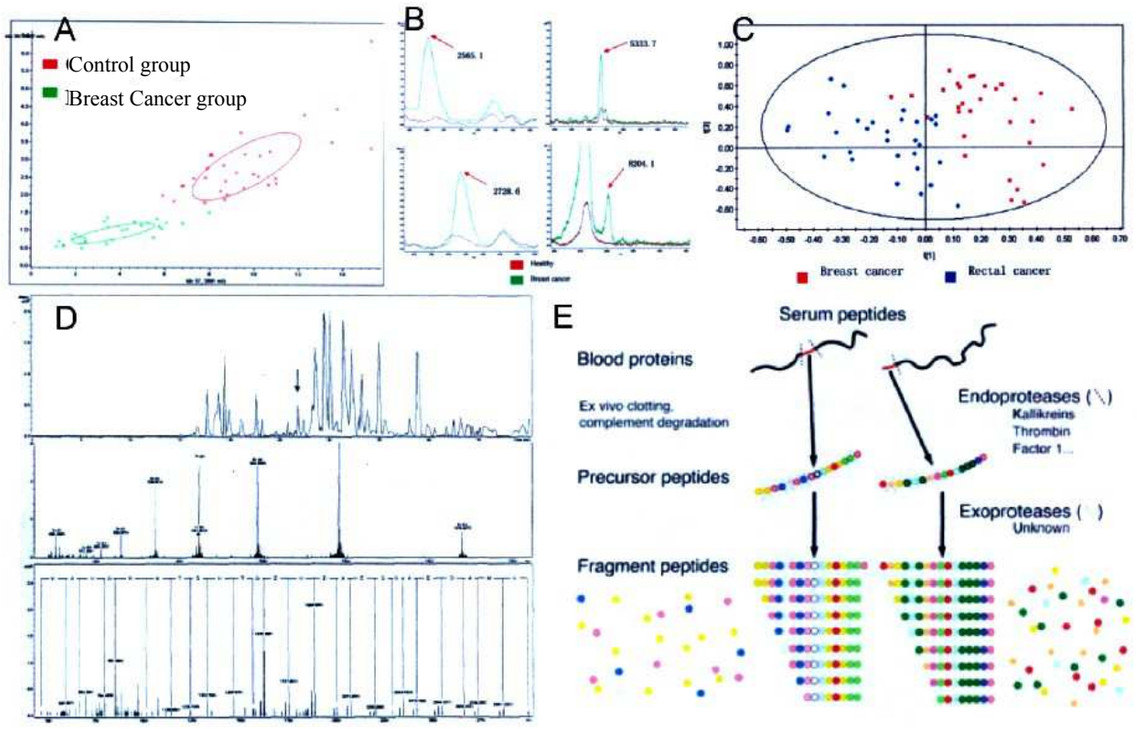


Fig. 8. Experimental Design of ClinTOF System for Breast Cancer Research



A: sample distribution diagram; B: average profiling of significantly differential peaks; C: distribution of samples obtained by SIMCAP+; D: results of MS identificdatin of differential polypeptides;
E: description of mechanism according to distribution of identified polypeptides

Fig. 9. Findings of Using the ClinTOF System to Breast Cancer Research

Some researchers have also shown that protein MS peaks obtained by the SELDI system can effectively screen and distinguish the breast cancer and non-breast cancer patients. Van Winden et al. adopted the SELDI technology to determine the density difference between the breast cancer group and the control group, making new progress compared with the previous breast cancer marker protein reseraches (van Winden et al, 2009). Gast et al. utilized the SELDI technology to detect the protein profiling in tissues and serum to diagnose breast cancer, showing that 3 peaks are significantly correlated to breast cancer. 27 tissues were detected to have differential peaks. These protein fragments presented potential pathological and physiological mechanisms related to breast cancer, which is helpful for raising the diagnosis rate of the breast cancer (Gast et al., 2009). Taku et al employed the SELDIS technology to detect 65 breast cancer patients as a group, concluding that the overexpression of a non-identified protein and the underexpression of the other are related to the lymphatic metastasis of breast cancer (Taku et al., 2006).

3.1.3 Nasopharyngeal carcinoma

In the early stage, if nasopharyngeal carcinoma fails to show some disease features due to disturbance of some factors, the test method is so insensitive that the patient cannot get timely and correct diagnosis, it will delay the treatment and cause unfavorable prognosis. In the research on nasopharyngeal carcinoma serum peptidome, some progress has been made.

We employed the ClinTOF system to research the marker of early-stage nasopharyngeal carcinoma. The magnetic bead system was SPE-C. By comparing 40 nasopharyngeal carcinoma cases and 61 healthy people, 65 significantly differential peaks were observed in the mass range of 1,000Da-10,000Da. The peak with the most obvious difference of expression has the mass-to-charge ratio of 1262.67 and is identified with the LTQ -Orbitrap to be the fragment of fibrinopeptide A. Fibrinogen is a kind of glycoprotein rich in blood and a symmetrical dimer composed of 6 polypeptide chains; it plays a major role in blood coagulation and hemostasis. So far, more than 300 fibrinogen natural mutants have been discovered, among which, about 55% are subclinical, 25% have hemorrhagic tendency and 20% have thrombophilia. Among these mutants, the most common expression is that one amino acid is replaced by another. At the molecular biology level, expressions are point mutation, deletion, insertion and nonsense mutation. Moreover, the polymorphism of the fibrinogen results in the overexpression, which raises the fibrinogen level of plasma and closely relates to diseases.

Researchers have also indicated that this system has a higher repeatability than the IMAC Cu^{2+} chips (Ciphergen Company) and is capable of finding more differential peaks in proteins with the molecular weight ranging 2,000Da -5,000Da.

Wei et al., using the SELDI system, screened out the mass-to-charge ratios of four proteins, to be detailed, 4 097Da, 4 180 Da, 5 912Da and 8 295Da, which make a marker combination to build a classification tree model for diagnosis of nasopharyngeal carcinoma. In diagnosing nasopharyngeal carcinoma, the sensitivity and specificity of this model were 94.5 % and 96.7% respectively and for the blind screening group, the two figures were 92% and 92.9% respectively. Using the protein MS peaks with the mass-to-charge ratio of 4 581Da and 7802Da to predict the stage I and stage II nasopharyngeal carcinoma, the accuracy proved to be 80% and for the stage III and stage IV, this figure was 86% (Wei et al, 2008).

Guo et al. applied the SELDI system and the artificial neural network technology to analyze serum of 58 type-A (cranial nerve type: featured by basicranial destruction and cranial nerve violation, no lymphonodi cervicales metastasis) and type-D (lymphonodi cervicales metastasis type: widespread metastasis of one- or two-sided lymphonodi cervicales, no cranial nerve violation or basicranial bone destruction) nasopharyngeal carcinoma patients and obtained 11 potential biomarkers, with the mass-to-charge ratio of 4 053, 5 885, 4 072, 5 798, 4 209, 8 689, 2 382, 9 357, 2 221, 4 230Da and 5 901Da respectively. The MS peaks of these proteins saw accuracy for distinguishing type A and type D reach 90% (Guo et al, 2005).

Cho et al. employed IMAC3-Cu protein chips to detect the serum of the healthy people, modified nasopharyngeal carcinoma patients after treatment and the relapse nasopharyngeal carcinoma patients before and after chemotherapy, discovering 13 meaningful protein MS peaks. Protein MS peaks with the mass-to-charge ratio of 2950Da and 6 701Da were selected to build a classification tree model for predicting chemotherapy response, with the sensitivity and specificity of 80% and 87% respectively. In detection and analysis of serum protein profiling of the healthy people and the relapse nasopharyngeal carcinoma patients, it is newly discovered that the MS peaks with the mass-to-charge ratio of 3803Da and 3953Da were found in the serum of both patients taking the initial diagnosis and treatment, and the modified nasopharyngeal carcinoma patients after treatment. In addition, the protein MS peaks with the mass-to-charge ratio of 3953Da and 7765Da were identified to be the inter-alpha trypsin inhibitor heavy chain H4 precursor (ITIH4) fragments and platelet factor 4 (PF4) with the method of MS/MS sequence and immune affinity capture test, thus presuming that they are related to the onset, development, metastasis and relapse of nasopharyngeal carcinoma (Cho et al., 2004).

Huang et al. adopted the CM 10 chip and SELDI technology to detect squamous epithelial cells of the nasopharyngeal carcinoma patients and made contrast with serum of the healthy people. 94 protein MS peaks were detected, among which, 26 were obviously different from the serum of healthy people ($P<0.05$, Mean>SD), 5 of overexpression and 21 underexpression. Three protein MS peaks with the mass-to-charge ratio of 3159 83, 5 187 65 and 1 3738.6Da were selected to build a diagnosis model. Verified by the blind method, this diagnosis model has accuracy of 90.63%, sensitivity 95.00%, specificity 83.33%, positive predicted value 90.48% and negative predicted value 90.90%, indicating that this model has a higher diagnosis value for NPC (Huang et al, 2008).

Doustjalali adopted the ClinProt system and the MASCOT database retrieval, discovering that in the serum of nasopharyngeal carcinoma patients, the content of CPL showed overexpression compared with the control group; further, after the ELISA method validation, it was discovered that in the nasopharyngeal carcinoma tissues, CPL showed overexpression. After 6 months' treatment, the CPL content in tissues reduced. CPL is a kind of copper-bearing glycoprotein and considered to be a key molecule for activating angiogenesis factor, which can promote growth of tumors, as shown by studies. This research monitored the changes of ceruloplasmin in company with the ease of the disease with the proteome technology, which would play a role in diagnosis, curative effect observation and prognoses evaluation, observation and monitoring of nasopharyngeal carcinoma (Doustjalali et al., 2006).

Liao et al. adopted 2-DE /MALDI-TOF-MS technology to screen tumor specific antigens of nasopharyngeal carcinoma. High-abundance protein elimination and desalting pretreatment

were performed over serum of the nasopharyngeal carcinoma metastasis group, nasopharyngeal carcinoma non-metastasis group and the control group. Then, analysis was made to the 3 groups of serum profiling, obtained 29 differential protein spots and identified 23 proteins. Through comparison of the cancer profiling and the profiling of the control group, it is discovered that the transferrin, zinc finger protein 544, thyroid hormone binding protein, NM 23H 1 protein and FAD synthetase showed underexpression in nasopharyngeal carcinoma patients. Yet lipoxygenase (LOX), serum amyloid protein A1, cytochrome P450, sICAM1, cathepsin G and histone lysine specific demethylase 1 showed overexpression in the two groups of nasopharyngeal carcinoma patients; the LOX, sICAM1, cathepsin G and histone lysine specific demethylase 1 showed higher expression in the nasopharyngeal carcinoma metastasis group than in the non-metastasis group, while the heat shock protein 70 was only expressed in the nasopharyngeal carcinoma metastasis group. In addition, by combining the immune histochemical technique and ELISA method, it is concluded that HSP70, sICAM1 and serum amyloid protein A (SAA) were potential serological markers for mediating nasopharyngeal carcinoma metastasis and also exerted an important role in clinical detection and control of nasopharyngeal carcinoma (Liao et al, 2008).

3.1.4 Cervical carcinoma

The serum peptidomics technology has been used in the research of endometrial carcinoma and is of guidance significance for the early diagnosis and clinical grading of diseases.

Our researches using the ClinTOF system have indicated that in the cervical carcinoma group and the control group, the protein peak intensify difference of only 21 proteins had statistical significance ($P<0.05$). M1450.35Da, M1778.7Da, M1896.65Da and M5520.42Da were taken as the classified variables to build a classification predictive model, to make classification diagnosis over the cervical carcinoma group and the control group, with the identification rate and predictive ability of 90.45% and 81.75% respectively. By contrasting the proteome of different pathological differentiation degrees, it is discovered that the protein peak intensity difference of 2 proteins had statistical significance ($P<0.05$). M5904.14Da and M5264.26Da were taken as the classified variables to build a classification model, to make classification diagnosis on different differentiation degrees, with the identification rate and predictive ability of 81.48% and 78.89% respectively.

Yoshizaki et al. adopted the SELDI system to analyze the protein profiling of 19 endometrial carcinoma cases and 20 normal tunica intima tissues, discovering two differential proteins, namely EC1 and EC2. The former had overexpression in the endometrial carcinoma tissues and the latter on the contrary. These differential proteins may be used for diagnosing endometrial carcinoma hopefully (Yoshizaki et al., 2005).

3.1.5 Leukemia

The treatment of acute leukemia (AL) depends on minimal residual disease (MRD) diagnosis and detection. However there has been no serum biomarker that can be used by the clinicians for diagnosing AL and evaluating MRD. Yang et al. analyzed serum of AL patients, discovering two peptides (m/z of 1778 and 1865) reduced with the rise in the modification degree and that 1865 was related to AL types. After further FT-ICR-MS

detection, the two peptides are the C3F fragment. The linear regression analysis has indicated that the combined use of the peptides could distinguish PML / RARα positive molecules from M3. Findings show that the two C3F fragments are significantly correlated with the MRD level and can be used for evaluating clinical MRD (Liang et al., 2010).

3.1.6 Brain glioma

Our forward-looking researches on brain spongiocytoma diagnosis have indicated: using the SPE-C magnetic bead separation and liquid automatic processing robot operations of the ClinTOF system, after MS, 74 differential peaks occurred in the serum of 84 patients and 72 healthy people. 55 samples were taken randomly to build the MS model of patients and healthy people, with the diagnosis specificity for the healthy people reaching as high as 96.4% (Fig.10).

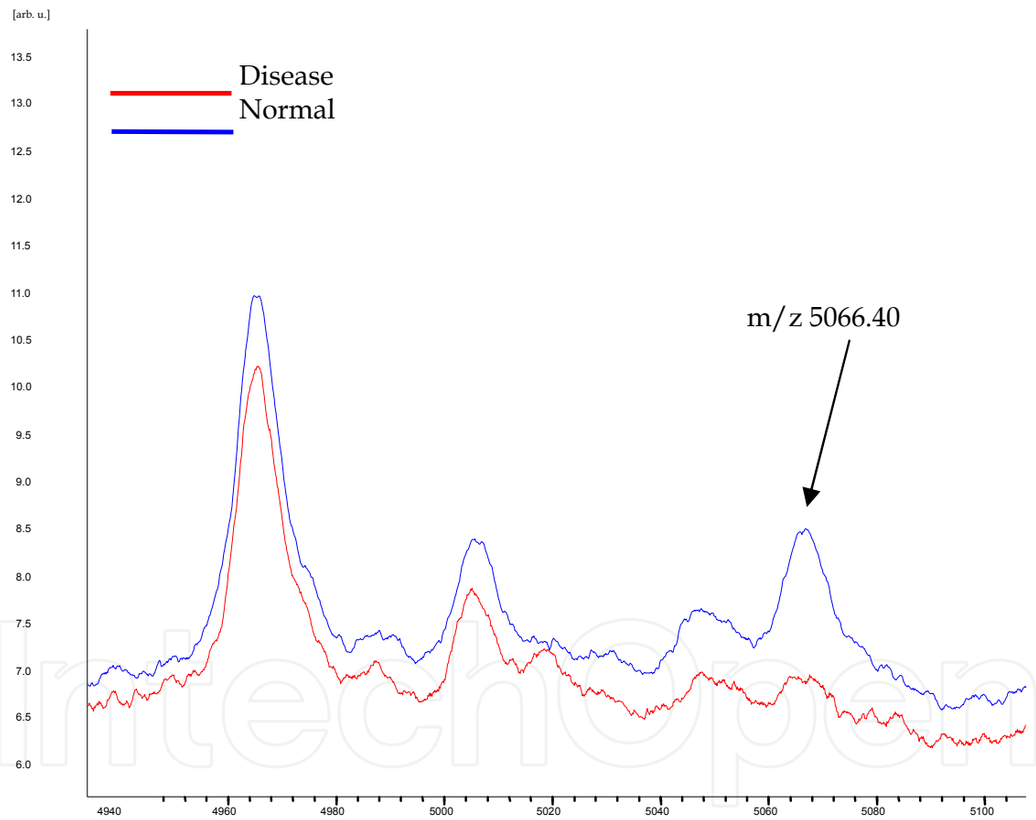


Fig. 10. Differential Polypeptide Peaks of Brain Glioma

3.2 Nerve and mental diseases

Alzheimer disease (AD) is a degenerative disease with unclear causes and mainly disturbing cerebral cortex neurons. Zhang et al. discovered that in the serum of AD patients, the haptoglobin and complement factor H were overexpressed. Haptoglobin is a kind of acute phase reactive protein and an antioxidant, featuring multiple biological activities and is considered to have participated in the inflammation process together

with lymphocyte, neutrophilic granulocyte and monocyte and played a vital role in regulating host defense. Complement factor H is an important complement modulation substance and can also work as a significant effector molecule of congenital immunity. It is capable of distinguishing self and non-self and the activated and non-activated cell surfaces. In the research, the overexpression of two proteins indicates the inflammatory reaction, oxidation stress and immunologic mechanism are participating in the onset process of AD (Zhang et al., 2004). Hye et al. made case-control study to the plasma of 50 AD patients and the control group, discovering many disease pathology related proteins, including complement factor H (CFH) precursor and α22 macroglobulin and pointing out that the expression of CFH is related to the cognitive dysfunction of AD patients. α22 macroglobulin (A2M) is a pan-protease inhibitor. In case of brain damages, the immunological activity of A2M of the neuron and colloid cells will rise, especially at the levels of senile plaques, nerve fiber matting and auantic neuron neurite (Hye et al., 2006).

We utilized the SPE-C magnetic beads and ClinTOF system to compare 85 two-way patients having no medicine and 100 healthy people, discovering 5 significantly differential peaks in the mass range of 800Da-12,000Da among the mental disease patients and the healthy people. The serum polypeptide MS standard model profiling has a predictive sensitivity of 100% for the healthy people and for the mental disease patients, 87%.

3.3 Healthy people screening

To study relevancy between some disease and polypeptide, we introduced serum polypeptide profiling into health screening. The ClinTOF system was used to study 1980 healthy people related polypeptide profiling researches, covering the age of 18-75, large sample statistics of both sexes, with the experimental design work shown in Fig.11. With

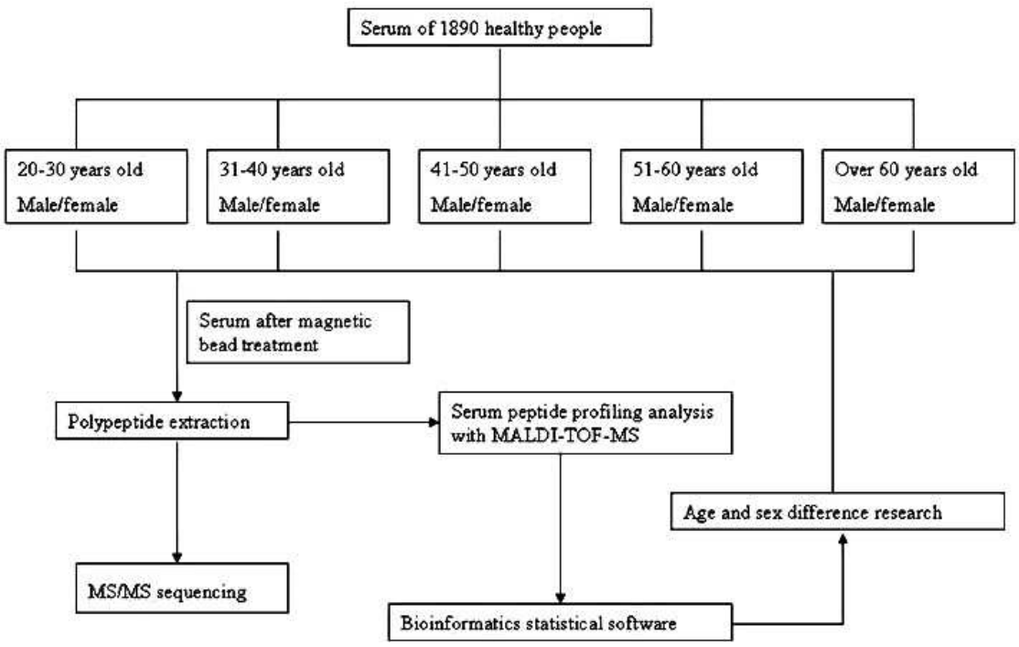
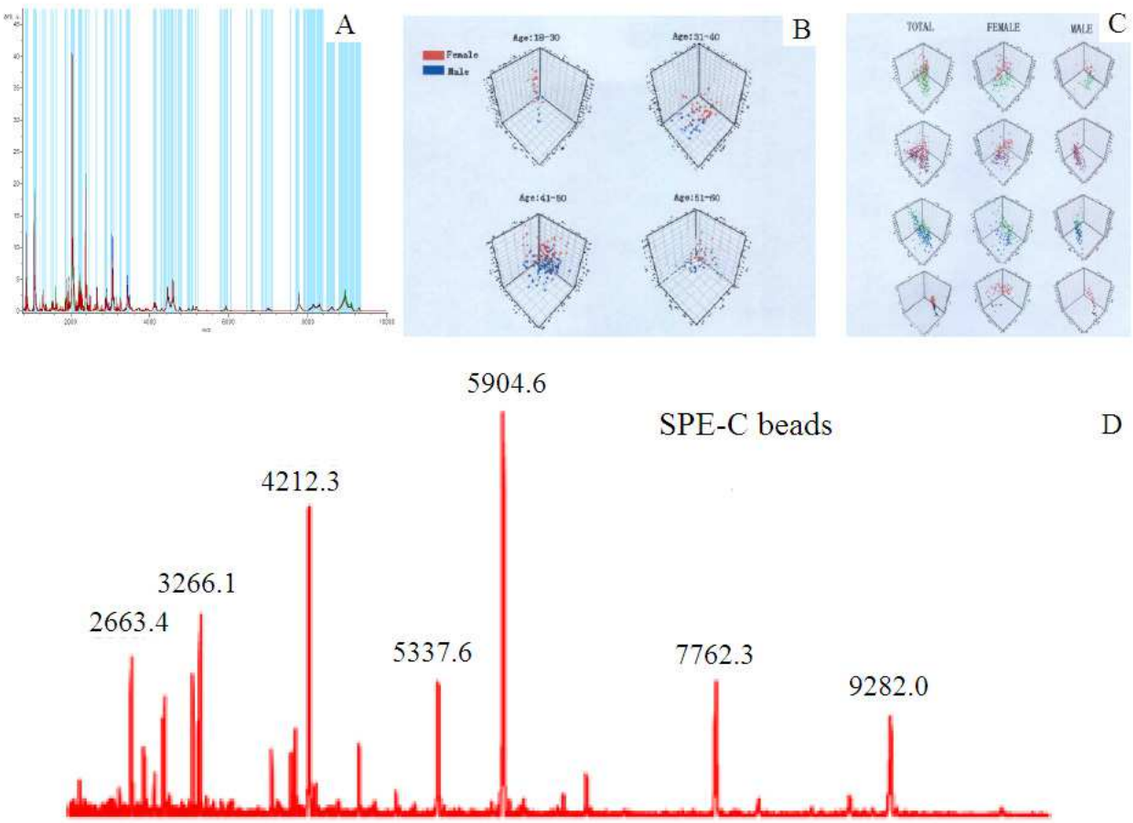


Fig. 11. Experimental Design of Healthy People Screening with the ClinTOF System

the SPE-C magnetic bead, MS profiling (Fig.12) for age 5 ranges was built and 10 peaks with the most significantly mass-to-charge ratio were identified. The online search of database with Mascot indicated that most polypeptide protein precursors were isogenous with the human prokinin precursors. After the analysis of serum peptidome profiling and bioinformatics data mining of 1890 healthy people, for the healthy people, the profiling has no obvious difference regardless of age and sex, but for people younger than 30, the sex difference factor shall be considered, that is, the elder the person is, the more different the serum peptidome profiling will be. Therefore, in researching the disease peptidome profiling, the age and sex matching factors shall be considered. It has also discovered that 50 polypeptides have a higher frequency of occurrence (more than 30%) in the serum of healthy people, indicating that these polypeptides can work as reference for health controls.



A: average profiling of 1890 healthy people; B and C: results from PCA analysis by age and sex; D: some peaks of highly-frequent occurrence detected by MS after the treatment by SPE-C magnetic beads

Fig. 12. Results of Healthy People Screening with the ClinTOF System

3.4 Fetal congenital aplasia

We have used the ClinTOF system to study fetal congenital aplasia, with the research protocol shown in Fig.13. This thinking has generalized the sample collection strategies and the result expectations sufficiently indicating the diversification and universality of serum peptide method, and can be used for most research areas. Secretion leukoprotease inhibitory factor (SLPI) is an endogenic immunity related protein, capable of specifically inhibiting elastolytic enzyme, cathepsin G, trypsin and fibrinolysin and boasts of antiphlogistic response and antibiosis/antiviral activity. The detection of the content of SLPI protein in serum may have great clinical significance for diagnosis. Differential polypeptides were obtained (Fig.14) and their functions were explained (Table 4).

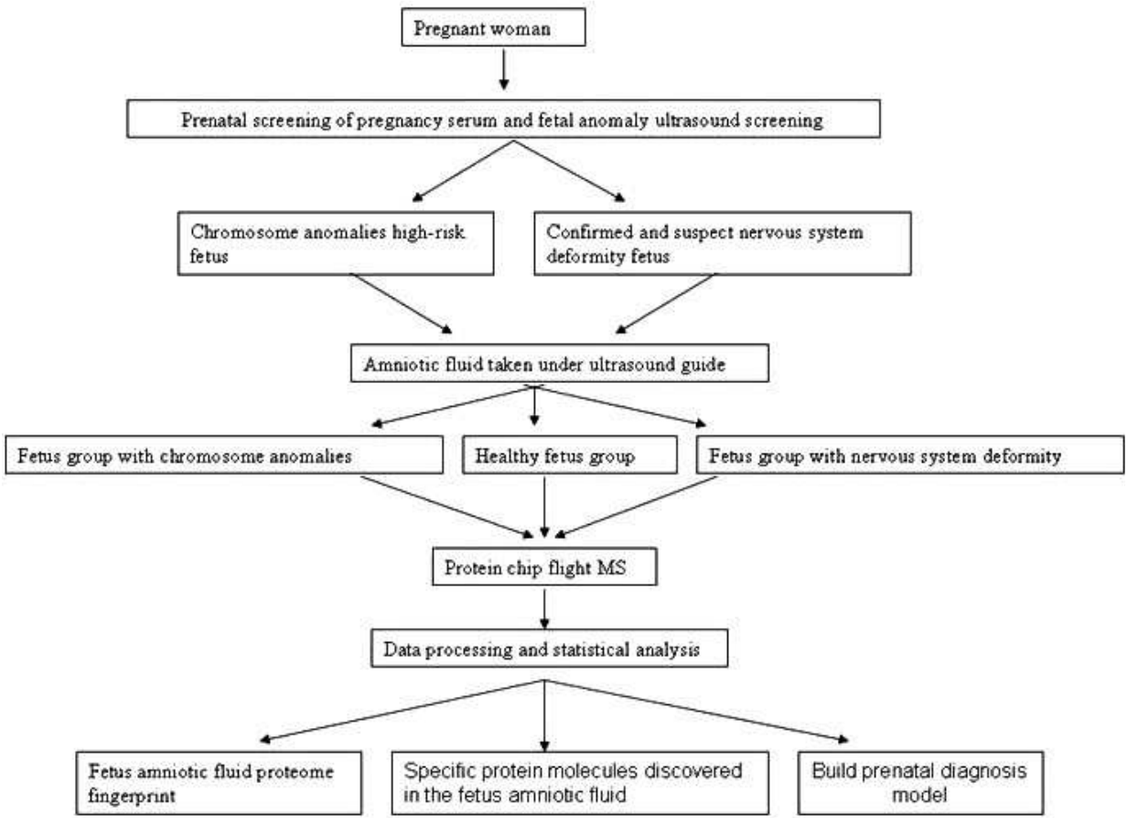
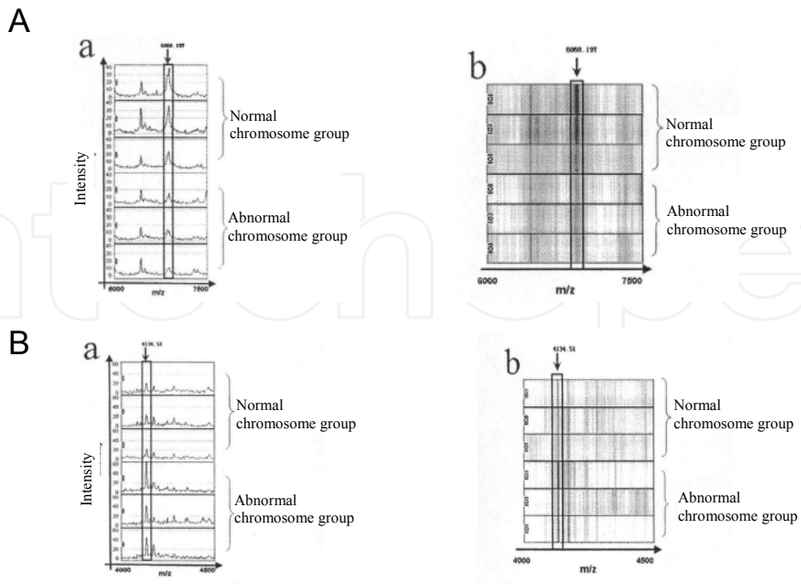


Fig. 13. Research Protocol for **Fetal Congenital Aplasia** with ClinTOF System



A and B: gel image of differential polypeptides

Fig. 14. Findings of Rsearching **Fetal Congenital Aplasia** with the ClinTOF System

biomarkers	Meaningful matching information in Swiss-Prot database			
				comments
(Mw)	Protein Name	Mw	pI	
3107.1	GLCM1_HUMAN(Q8IVK1) Putative glycosylation-dependent cell adhesion molecule 1.	3109	8.23	ALTERNATIVE PRODUCTS
4967.526	HMHB1_HUMAN (O97980) histocompatibility protein HB-1	4965	4.95	Precursor of the histocomplatibility Minor antigen HB-1
5589.2	CO031_HUMAN (Q13653) Putative uncharacterized protein C15orf31.	5583	7.59	Expressed at higher level in total thymocytes than in mature T or NK cells， especially in CD8+ cells.
11717.0	LV201_HUMAN (P01704) Ig lambda chain V-II region TOG	11713	8.66	MISCELLANEOUS: This is a Bence-Jones protein. SIMILARITY: Contains 1 Ig-like (immunoglobulin-like) domain.
	HV308_HUMAN (P01769)	11706	9.80	MISCELLANEOUS: This chain was isolated from a Waldenstrom's macroglobulin. SIMILARITY: Contains 1 Ig-like (immunoglobulin-like) domain.
	SLPI_HUMAN (P03973) Antileukoproteinase	11726	9.10	Acid-stable proteinase inhibitor with strong affinities for trypsin， chymotrypsin， elastase， and cathepsin G. May prevent elastase-mediated damage to oral and possibly other mucosal tissues.

Table 4. Retrieval Results from Swiss-Prot Protein Database

Besides diseases studied by the ClinTOF system above, for the moment, we have finished the detection of near 10000 samples. Diseases are not only restricted to tumors, but covering almost all clinical aspects, as shown in Table 5.

Disease	Case Load	Source
Lung cancer, lung squamous carcinoma, adenocarcinoma of lung	1678	PLA 301 Hospital, Guangzhou People's Hospital, Xian Jiaotong University, Hebei Medical University Fourth Hospital
Stomach cancer	117	Fudan University, Shanghai Ruijin Hospital
Esophagus cancer	48	West China University of Medical Science, CICAMS
Nasopharyngeal carcinoma	238	Sun Yat-Sen University Cancer Center
Breast cancer	156	Beijing Cancer Hospital, Xinjiang Medical University
Pancreatic cancer	120	Guangzhou People's Hospital
Endometrial carcinoma Endometrial carcinoma	283	Beijing Obstetrics and Gynecology Hospital, Capital Medical University
Leukemia	103	Shanghai Fifth Hospital
Intestinal cancer	185	Shanghai Jiaotong University, Capital Medical University
Rheumatosis, lupus erythematosus	75	Shenzhen People's Hospital
Cirrhosis, hepatic carcinoma, hepatitis B	303	PLA 301 Hospital, Tianjin Third Hospital
Cognitive dysfunction	230	Zhongnan Hospital of Wuhan University, Shanghai Jiaotong University
Endometriosis uterine	137	Beijing People's Hospital
Diabetes	188	Shanghai Changzheng Hospital
Test-tube baby tracking	102	Beijing University Third Hospital
Mental disease	1220	Chongqing Medical University
Healthy people	2473	Capital Medical University
Athletes	1995	PLA 301 Hospital

Table 5. Diseases and Case Load Applied by the ClinTOF System (alterable)

4. Prospect

Clinically, the high incidence and mortality rate of complex diseases like tumors cry for the new early-diagnosis patterns and effective early-diagnosis markers. In the mining of the mass spectrometric data, the cluster analysis and classification analysis have been used, but many problems still need further studies. In particular, the pretreatment of MS profiling, there is no standard method available. These issues are critical for putting serum peptidome into clinical detection. Secondly, along with the instant development and application of

mass spectrometer and serum peptidome technology, a lot of serum polypeptide profiling data have been produced. It has become an international competition focus in the post-genome era how to develop an effective bioinformatics instrument to determine proteins related to some vital phenomena (like growth, upgrowth, tumor onset) from the substantive data including peptide mass and intensity and their functions. It is believed that with the progressive development and application of serum peptidomics technology and the ClinTOF system, better clinical models will appear and more early-stage tumor biomarkers will be discovered and identified. Also, great breakthroughs will be made in the pathogeny, diagnosis and treatment of complex diseases, to cultivate new hope for research, diagnosis and treatment of many diseases.

5. References

[1] Andrew Tarnaris, Laurence D Watkins and Neil D Kitchen. Biomarkers in chronic adult hydroce-phalus. Cerebrospinal Fluid Research, 2006,3:11.

[2] Brichory F, Beer D, Le Naour F, et al. Proteomics- based identification of protein gene product as a tumor antigen that induces a humoral immune response in lung cancer. *Cancer Res*, 2001, 61(21) : 7908- 7912.

[3] Cho WC, Yip TT, Yi PC, et al. Identification of serum amyloid a protein as a potentially useful biomarker to monitor relapse of nasopharyng eal cancer by serum proteomic profiling. *Clin Cancer Res*, 2004, 10: 43-52.

[4] Cruz-Marcelo A, Guerra R, Vannucci M, et al. Comparison of algorithms for pre-processing of SELDI-TOF mass spectrometry data. *Bioinformatics*, 2008, 24:2129-2136.

[5] Cubizolles M, Laurendeau I, Bedossa P. Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases. *Hepatology,* 2005; 41: 40-47.

[6] Doustjalali SR, Yuso f R, Gov indasamy GK, et al. Patients with nasopharyngea carcinoma demonstrate enhanced serum and tissue cerulo plasm in expression. *J Med Invest*, 2006, 53: 20-28.

[7] Gast M C, van Du lken E J, van Loenen TK, et al. Detection of breast cancer by surface enhanced laser desorption/ionization time of flight mass spectrometry tissue and serum protein profiling. *Int J BioMarkers*, 2009, 24 (3): 130 -141.

[8] Guo X, Cao SM, Y u JK, et al. Distinct seruma proteomic patterns between ascending and descending types of locoregionally advanced nasopharyng eal carcinoma assessed by surface enhanced laser desorption ionization and analyses. *Chin Med J*, 2005, 118(22): 1912-1917.

[9] Huang Y J, Xuan C, Zhang BB, e t al. SELDI-TOF-MS profiling of serum for detection of nasopharyngea carcinoma.*Cancer*, 2008 , 112(3): 544-551.

[10] Hye A , Lynham S, Thambisetty M , et al. Proteome based plasma biomarkers for Alzheimer ’s disease. *Brain* , 2006,129 (Pt11): 3042-3050.

[11] Liao QL, Zhao L, Chen X, et al. Serum proteome analysis for profiling protein markers associated with carcinogenesis and lymph nodemetastasis in nasopharyngea l carcinoma. *Clinical and Experimental Metastasis*, 2008, 25(4): 465-476.

[12] Paradis V, Degos F, Dargère D, Pham N, Belghiti J, Degott C, Janeau JL, Bezeaud A, Delforge D, Poon TC, Yip TT, Chan AT, Yip C, Yip V, Mok TS, Lee CC, Leung TW, Ho SK, Johnson PJ. Comprehensive proteomic profiling identifies serum proteomic signatures for detection of hepatocellular carcinoma and its subtypes. *Clin Chem,* 2003, 49: 752-760.

[13] Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC et al: Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, 2002, 359(9306):572-577.

[14] Taku N, Sharon KH , Steve RM, et al. Proteomic profiling of primary breast cancer predicts axillary lymph node metastasis. *Cancer Res*, 2 0 0 6, 6 6 (2 4) : 1 1 82 5 - 1 1 8 3 0.

[15] van Winden AW, Gast MC, Beijnen JH, Rutgers EJ, Grobbee DE, Peeters PH, van Gils CH. Validation of previously identified serum biomarkers for breast cancer with SELDI-TOF MS: a case control study. *BMC Med Genomics*,2009,2:4.

[16] Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, DeNoyer L, Fleisher M, Robbins RJ, Tempst P. Correcting common errors in identifying cancer-specific serum peptide signatures. *J Proteome Res*,2005, 4(4):1060-1072.

[17] Wasinger VC, Cordwell S J, Cerpa PA, et al. Progress with gene product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis, 1995, 16 (7): 1090-1094.

[18] Wei YS, Zheng YH, Zhang ZH, et al. Identification of serum biomarkers for nasopharyngea carcinoma by proteomic analysis. *Cancer*, 2008, 12 (3): 544-551.

[19] Yoshizaki T, Enomoto T, Nakashima R, e t al. Altered protein expression in endometrial carcinogenesis. *Cancer Lett*, 2005, 226 (2):1012 - 1061.

[20] Zhang R, Barker L, Pinchev D, et al. Mining biomarkers in human sera using proteomic tools. *Proteomics*, 2004, 4(1): 244-256.

[21] Johnson, KL, Mason, CJ, Muddiman, DC, etal. Analysis of the low molecular weight fraetion of serum by LC-Dual ESI-FT-ICR mass speetrometry:spectrometry: Precision of time, mass, and ion abundance. Anal Chem, 2004, 76:5097-5103.

[22] Soltys SG, Le QT, Shi G, etal. The Use of plasma Surface-Enhanced Laser Desorption/ionization Time-of-Flight Mass Spectrometry proteome patterns for detection of head and neck squamous cell cancers, Clinical Cancer Research, 2004, 10:4806-812.

[23] Honda K, Hayashida Y, Umaki T, et al.Possible detection of pancreatic cancer by plasma protein profiling. Cancer Res, 2005, 65：10613-10622.

[24] Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activity confer tumor-specific serum Peptidome Patterns. J Clin Invest, 2006, 116:271-284.

[25] Novak K. Biomarkers:Taking out the Trash. Nature, 2006, 6:92.

[26] Cheng AJ, Chen LC, Chien KY, et al. Oral Cancer Plasma Tumor Marker identified with Bead-Based Affinity Fractionated Proteomic Technology. Clin Chem, 2005, 51:2236-2244.

[27] YIP T TC, Chan J WM, Cho W CS, et al. Protein chip array profiling analysis in patients with severe acute respiratory syndrome identified serum amyloid: a protein as a a biomarker potentially useful in monitoring the extent of pneumonia. Clin Chem, 2005, 51(l):47-55.

[28] Semmes OJ, Feng Z, Adam BL, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of Prostate cancer: assessment of platform reproducibility. Clin Chem,, 2005, 51:102-12.

[29] Dekker LJ, Boogerd W, Stockhammer G, Et al. MALDI-TOF mass spectrometry analysis of cerebrospinal fluid tryptic peptide profiles to diagnose leptomeningeal metastases in patients with breast cancerMol Cell Proteomics, 2005, 4:1341-1349.

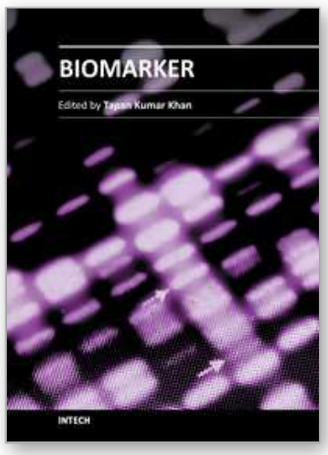
[30] Villanueva J, Martorella AJ, Lawlor K, et al. Serum peptidome patterns that distinguish metastatic thyroid carcinoma from cancer-free controls are unbiased by gender and age. Mol Cell Proteomic, 2006, 5(10):1840-1852.

[31] Dobrin N, Urban A.K, Erie E.N., et al. Population Proteomics. Mol Cell Proteomics 2006, 5(10):1811-1818

[32] Agranoff D, Delmiro FR, Papadopoulos MC, et al.. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. Lancet, 2006,368(9540):1012-1021

[33] Emanuel FP, Claudio B, Robyn P, et al. The blood peptidome: a higher dimension of information content for cancer biomarker discoverNat Rev Cancer,2006,6(12):961-967

IntechOpen



Biomarker

Edited by Prof. Tapan Khan

ISBN 978-953-51-0577-0
Hard cover, 392 pages
Publisher InTech
Published online 27, April, 2012
Published in print edition April, 2012

Clinicians, scientists, and health care professionals use biomarkers or biological markers as a measure of a person's present health condition or response to interventions. An ideal -biomarker should have the following criteria: (I) ability to detect fundamental features of the disease, (II) ability to differentiate from other closely related diseases, (III) ability to detect early stages and stages of progression, (IV) the method should be highly reliable, easy to perform and inexpensive, and (V) sample sources should be easily accessible from body. Most of the chapters in this book follow the basic principle of biomarkers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Kaihua Wei, Qingwei Ma, Yunbo Sun, Xiaoming Zhou, Weirong Guo and Jian Yuan (2012). Serum Peptidomics, Biomarker, Prof. Tapan Khan (Ed.), ISBN: 978-953-51-0577-0, InTech, Available from: http://www.intechopen.com/books/biomarker/serum-peptidomics

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen