SCIENCE CHINA
Life Sciences

• REVIEW •

December 2014 Vol.57 No.12: 1162–1171
doi: 10.1007/s11427-014-4714-2

Rapid development of proteomics in China: from the perspective of the Human Liver Proteome Project and technology development

LI Ning[1], XU ZhongWei[1], ZHAI LinHui[1], LI YanChang[1], FAN FengXu[1, 2], ZHENG JunJie[1], XU Ping[1*] & HE FuChu[1*]

[1]*State Key Laboratory of Proteomics, National Engineering Research Center for Protein Drugs, Beijing Proteome Research Center, National Center for Protein Sciences Beijing, Beijing Institute of Radiation Medicine, Beijing 102206, China*
[2]*Graduate School of Anhui Medical University, Hefei 230032, China*

Received April 14, 2014; accepted July 1, 2014; published online August 12, 2014

Proteomics focuses on the systematic identification and quantification of entire proteomes and interpretation of proteins' biological functions. During the last decade, proteomics in China has grown much faster than other research fields in life sciences. At the beginning of the second decade of the 21st century, the rapid development of high-resolution and high-speed mass spectrometry makes proteomics a powerful tool to study the mechanisms underlying physiological/pathological processes in organisms. This article provides a brief overview of proteomics technology development and representative scientific progress of the Human Liver Proteome Project in China over the past three years.

proteomics, LC-MS/MS, HLPP, protein-protein interaction, protein posttranslational modification, bioinformatics

Citation: Li N, Xu ZW, Zhai LH, Li YC, Fan FX, Zheng JJ, Xu P, He FC. Rapid development of proteomics in China: from the perspective of the Human Liver Proteome Project and technology development. Sci China Life Sci, 2014, 57: 1162−1171, doi: 10.1007/s11427-014-4714-2

Proteomics investigation in China has been paced over 10 years under the guidance of the China Human Proteome Organization (CNHUPO), which was formally established in 2003. Under the consistent and high levels of funding support, proteomics study in China has achieved rapid development. In the last three years, He FuChu and colleagues had already reviewed the progress of proteomics study in China [1–4]. Here, we review the representative progress made by proteomics scientists in mainland China from 2010 to 2013. Concretely, we will mainly focus on the progress of the Human Liver Proteome Project (HLPP) and related technology development.

1 Achievements of the HLPP in China

He FuChu and colleagues in CNHUPO designed and advocated one road map for HLPP and HPP (Human Proteome Project) studies (i.e., two profiles: expression and modification; two maps: interaction and localization; three repositories: sample, antibody and data; and two outputs: physiology and pathology), which had been presented as early as the first HUPO Workshop for Human Proteome Initiatives (Bethesda, MD, USA, April 2002) and were included in the HLPP [5]. Under this guidance, Chinese proteomists have made great progress in the last three years (Figure 1).

1.1 Liver proteome expression profile

In the first pilot experiments, Chinese proteomists and in-

*Corresponding author (email: hefc@nic.bmi.ac.cn; xupingghy@gmail.com)

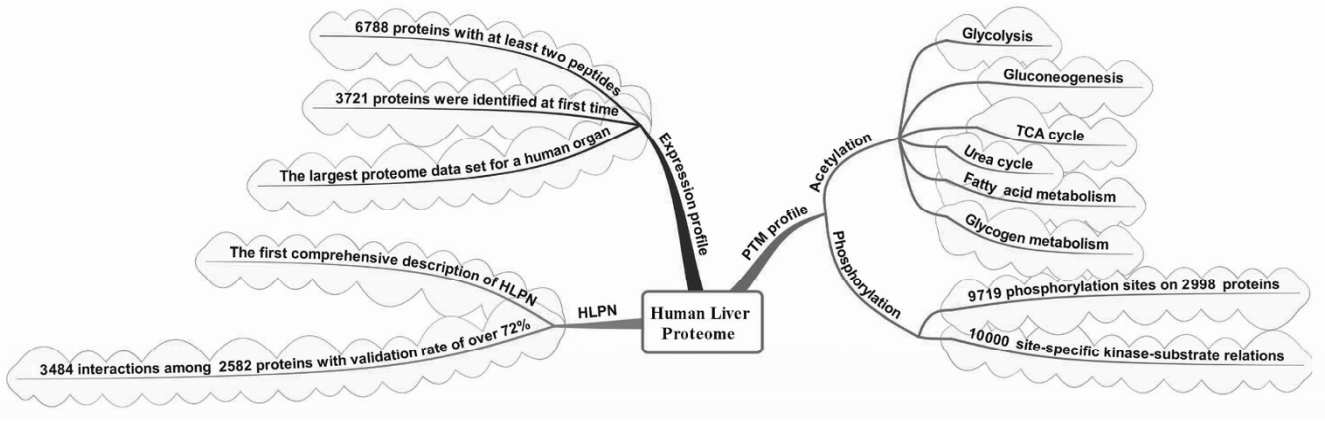

Figure 1 Major achievements of the Chinese proteomics community across the "solar system" of the HLPP. HLPN, human liver protein interaction network; PTM, post-translational modification.

ternational collaborators had made a huge effort and great progress on the Chinese liver proteome expression profile. After evaluation of individual variation in protein expression levels in normal liver, standard operating procedures (SOPs) for HLPP sample preparation were generated [6]. By collaborative and systematic analysis, a total of high confident 6788 proteins with at least two peptides matches were identified [7,8]. Of these, 3721 proteins were identified for the first time in the liver. This is the largest proteome data set for a human organ that has been published by any of the initiative projects in HUPO to date. Analysis of this data set showed that the HLP spans six orders of magnitude in relative protein abundance. It is also noticed that 78% of the proteins fell in the middle of this range. Of the 3721 newly identified liver proteins, 82.5% were of low abundance. Interestingly, four members of the cytochrome P450 family and three ion channels were found and described for the first time in this organ. These results validated the great effort the Chinese proteomists had put forth and the sensitivity of the platform they used for this project, especially considering that mass spectrometry (MS) was not so powerful several years ago. The *Journal of Proteome Research* published one special issue for the outcome of the HLPP and greatly acknowledged the contribution of this international project [7,8].

1.2 Liver proteome modification profile

The studies on the liver proteome modification profile have been implemented extensively for phosphorylation and acetylation. The groups of Xiong Yue, Guan KunLiang, and Zhao ShiMin at Fudan University utilized one set of high-efficiency antibodies they developed to enrich acetylated peptides from liver samples, and found that lysine acetylation involved in the intermediate metabolism is a prevalent modification form. Their finding indicated that almost all of the enzymes in glycolysis, gluconeogenesis, the tricarboxylic acid (TCA) cycle, urea cycle, fatty acid metabolism, and glycogen metabolism are acetylated in human liver tissue. They also showed that the acetylation status of metabolic enzymes was regulated by the concentration of metabolic fuels, such as glucose, amino acids, and fatty acids. Based on this study, they concluded that acetylation plays a major role in metabolic regulation [9]. Moreover, through collaboration with this group, Zhao GuoPing and Zhao ShiMin demonstrated that the metabolic regulatory mechanism modulated by acetylation is conserved from bacteria to mammals. In this signaling pathway, central metabolic enzymes in *Salmonella* were acetylated extensively and differentially as well, in response to different carbon sources, concomitant with changes in cell growth and metabolic flux. Acetylation also regulated the relative activities of rate-limiting enzymes, switching the direction of glycolysis versus gluconeogenesis or directing the branching between the citrate cycle and the glyoxylate cycle [10]. These studies convincingly demonstrated that the acetylation not only is common and conserved in cells, but also plays fundamental roles in regulating the metabolism process of organism.

Because of their outstanding achievement and great contribution to the proteomics field, Xiong Yue and Guan KunLiang were awarded the 2nd CNHUPO Academic Contribution Award at the 8th CNHUPO Conference in Chongqing.

Further, by integration of phosphoproteomic and computational analyses, Xue Yu and Zou HanFa experimentally identified 9719 phosphorylation sites on 2998 proteins in human liver. They further predicted a human liver protein phosphorylation network containing over 10000 site-specific kinase-substrate relationships [11].

1.3 Interaction and localization maps

Protein-protein interaction (PPI) profiling contributes to elucidate protein functions, metabolism and signaling pathways, molecular machines, and functional protein modules,

which are also one of the major goals of the HLPP plan. So far, proteome-scale protein interaction maps have been generated for many organisms, including bacteria, yeast, worms, and flies. These maps have provided substantial new insights into integrative biology, mechanisms of disease, and drug discovery. However, only a small fraction of the total number of human PPIs has been identified. Yang Xiao-Ming's and He FuChu's groups screened 5026 human liver proteins by yeast two-hybrid technology and established a human liver protein interaction network (HLPN) composed of 3484 interactions among 2582 proteins. Some of the identified proteins are metabolic enzymes that occur exclusively in the liver and contribute to the phenotype of the liver. Some are even liver-disease-related proteins. They also did bench experiments with three independent biochemical or cellular assays to verify the PPIs, and found that the validation rate was as high as 72%, indicating the high quality of their data. This work represents the first comprehensive description of a HLPN, which provides a valuable tool for understanding the function of the protein interaction network of the human liver [12].

2 Technology development in China

There is no doubt that development of technologies is critical in proteomics research. In the last three years, continuous improvements in sample preparation, optimization of chromatography systems, enrichment of post-translational modifications (PTMs), label quantitation, and bioinformatics were made by the Chinese proteomics community. The proteomics platforms have been set up at the international level with high throughput, high sensitivity, and high resolution. They were applied to large-scale and quantitative proteomics-based research to accelerate the pace of the biological studies.

2.1 Protein preparation and proteolysis

The major challenge in sample preparation for proteomics studies is the diversity of proteins with wide dynamic range. Multiple labs in China carried out extensive investigations to overcome these issues in the last three years. Yang PengYuan's and Zhang XiangMin's groups [13] synthesized magnetic mesoporous silica microspheres with immobilized copper(II) to selectively enrich hydrophobic and hydrophilic peptides from complex samples. Zou HanFa and his colleagues [14] synthesized yolk-shell magnetic mesoporous carbon microspheres for global enrichment of low-abundance peptides with high efficiency. Ding Chen et al. [15] at the Beijing Proteome Research Center (BPRC) developed a concatenated tandem array of transcription factor response elements in DNA to enrich activated transcription factors with low abundance from cells for high-efficiency profiling. Liu QiongMing et al. [16] from the same lab utilized a similar approach and successfully characterized low-abundance endogenous nuclear receptors from mouse liver .

Serum is an appropriate resource for biomarker studies. However, the enrichment of low-abundance proteins is challenging. Zou HanFa's lab synthesized a highly ordered mesoporous carbon material (OMC) using a soft-template method. By utilizing this OMC as an adsorbent, they extracted endogenous peptides from human serum with high efficiency, resulting in identification of 3402 different peptides from only 20 μL human serum [17]. Zhang YuKui and Chen LangXing's lab [20] developed Fe_3O_4 nanoparticles covalently functionalized with iminodiacetic acid (IDA)-Cu to deplete abundant proteins in blood samples for proteomics analysis. At the same time, Zhang YuKui and Zhang LiHua's lab [19] also prepared IDA-functionalized magnetic nanoparticles to selectively remove bovine hemoglobin. These materials and technology will greatly improve the power of biomarker screening in proteomics analysis.

Proteolysis is one of the most important steps in the preparation of peptides for successful proteomics analysis. Liu Yun et al. [20] developed a microchip reactor that they assembled by layer-by-layer electrostatic binding of poly (diallyldimethylammonium chloride) and gold nanoparticles. Using the new microchip reactor, in which the enzymes were immobilized on the large specific surface-to-volume ratio network, the maximum proteolytic rate of the adsorbed trypsin was 400 mmol L^{-1}/(min μg). They also found that the enzyme-coated microchip could be coupled with LC-MS to implement online protein identification. Zhang XueYang et al. [21] developed on-target aptamer immobilization and laser-accelerated proteolysis. In this approach, they also modified the target plate with a layer of gold. They used this novel aptamer to detect samples using a MALDI-TOF MS strategy, and found that high specificity and sensitivity were achieved using as little as 2 μL of sample. Yao GuoPing et al. employed laser irradiation (808 nm) to enhance enzyme digestion at the proteomics sample preparation step. They found that laser radiation accelerated both in-solution and in-gel digestion that it spends only seconds on the tryptic proteolysis. This is a straightforward, fast, efficient, and inexpensive approach to high-throughput proteome analysis [22]. Some other approaches have also been developed to prepare samples, such as tandem digestion, which could be used to improve the coverage for proteome sample analysis [23]; and trypsin immobilization on hairy polymer chains on the surface of hybrid magnetic nanoparticles, which can lead to ultra-fast, highly efficient proteome digestion [24]. Chen HeMei et al. [25] synthesized and applied a kind of core-shell-shell composite Fe_3O_4@SiO_2@PMMA microspheres for enrichment of samples. They found that low-concentration peptides and proteins could be enriched effectively, rapidly, and conveniently using these microspheres.

Furthermore, Zhang YuKui and Zhang LiHua's group [26] developed an online platform for automatic analysis of proteomics samples by integrating online sample buffer exchange, protein enrichment, and digestion. They showed that the performance of such an integrated sample treatment device was comparable to that of the traditional offline sample treatment method, but that the time consumed was significantly reduced to 0.5%, which is useful for high throughput proteomics studies.

More interestingly, Zou HanFa's group [27] found that trypsin has ligase activity in addition to its proteolysis function. This is the first time that trypsin has been used as a protease to catalyze the digestion of proteins and as a ligase to catalyze the linking of isotope-labeled amino acids to the N-termini of tryptic peptides for reliable and accurate quantitative proteomics analysis. This finding also suggests that the splicing variants identified previously may be called into question, and a more stringent method is necessary for this kind of study.

2.2 Technology development for analysis of PTM

PTM plays an important role in almost all biological processes. PTMs, such as phosphorylation, glycosylation, and ubiquitination, are involved in common mechanisms for controlling the behavior of proteins, and also increase the complexity of proteome. However, the sub-stoichio- metric abundance and transient expression of the modified proteins increase the dynamic range of the sample and the signal, which is usually suppressed by high-abundance proteins. Therefore, it is necessary to develop PTM enrichment methods for the MS analysis.

2.2.1 Technology development for phosphoproteomics

Protein phosphorylation is one of the most biologically relevant and ubiquitous PTMs, and is the subject of the most routine PTM analysis to date. Chinese proteomists also contributed extensively to the technology development for phosphopeptide enrichment, LC-MS analysis, bioinformatics support, and functional study.

In order to test the state of phosphorylation and sample quality, Zou HanFa and his colleagues [28] combined strong anion-exchange monolithic capillary liquid chromatography and MALDI-TOF MS for sensitive detection of phosphopeptides in protein digests. To improve the enrichment efficiency of phosphopeptides, they developed a specific method for enriching phosphopeptides by anion exchange followed by flow-through enrichment for comprehensive profiling of phosphopeptides [29].

To selectively capture phosphopeptides with high efficiency, multiple new materials, such as Ti-hexagonal mesoporous silica, hydrophilic polydopamine-coated graphene for metal ion immobilization, hierarchical Ti-aluminophosphate-5 molecular sieves, and adenosine functionalized metal immobilized magnetic nanoparticles have been developed as well [30–33]. To understand the competition of phosphopeptides during enrichment, Ji JianGuo's lab carefully investigated the ratio of peptide to TiO_2 beads. Based on these data, they developed a two-step enrichment protocol to separate multi- and mono-phosphorylated peptides with citric acid [34,35]. To further improve the coverage for the analysis of the phosphoproteome, Zou HanFa's group tested the idea of a tandem digestion approach for sample preparation. By combining the data set from the first Glu-C digestion and the second trypsin digestion, they identified 8062 unique phosphopeptides and 8507 phosphorylation sites in HeLa cells, which are almost double the amounts achieved with their conventional trypsin digestion approach [23]. After enrichment, Zou HanFa and his colleagues tested multidimensional separation of phosphopeptides before LC-MS analysis. They separated enriched phosphoproteome samples into a large number of fractions from the first dimensional reversed phase (RP) LC separation at high pH. Then they pooled every two fractions with equal time interval, one from the section eluted early and another one from the section eluted later. The pooled fractions were finally analyzed by RPLC at low pH coupled with tandem MS analysis. They found the resulting 2D separation was highly orthogonal and yielded more than 30% phosphopeptide identifications over the conventional RP-RPLC approach. Their result further suggested that phosphorylation is more abundant and diverse than current proteomics platform can accommodate [36].

Not only qualitative but also quantitative analysis of PTMs is important for interpreting their biological significance. Zou HanFa's lab utilized Ti(4+)-EPO nanoparticles as the adsorbent for *in situ* solid phase enrichment and isotopic labeling of endogenous phosphopeptides [37]. They also developed a pseudo-triplex stable isotope dimethyl labeling approach for high accuracy and throughput of comprehensive quantitative phosphoproteome analyses. These methodologies have great potential for application in high-throughput analyses of biological samples for screening and discovery of disease-specific biomarkers [38].

Besides, Zou HanFa's group [39] developed a classification filtering strategy to improve the coverage and sensitivity of phosphoproteome analysis by utilizing the differential dissociation behavior of different phosphopeptides. They made an in-house collected human serum pro-peptidome target/decoy database (HuSPep) to accelerate database searching and increase the sensitivity of phosphopeptide identification [40]. By utilizing large-scale phosphoproteome datasets, Li YiXue and his colleagues found that the phosphosites in the vertebrate-specific functional modules are more highly conserved than in the basic functional modules and their flanking regions. Based on these data, they suggested that phosphorylation may have played an essential role in the evolution of vertebrates [40]. Yao XueBiao and Xue Yu's lab analyzed single nucleotide polymorphism (SNP) and phosphosite information with

their kinase-specific phosphorylation site predictor (GPS 2.0), and found that approximately 70% of the reported nonsynonymous SNPs are potential phosSNPs. They also showed that approximately 74.6% of these potential phosSNPs might also induce changes in protein kinase types in adjacent phosphorylation sites rather than creating or removing phosphorylation sites directly. These results shed light on the functional implication of such genetic variations, which may aid in understanding genetic variation in the context of human diseases and personalized medicine [42].

2.2.2 Technology development for glycoproteomics

Because of biological significance, multiple proteomics labs in China, including but not limited to those of Yang Peng-Yuan, Qian XiaoHong, Zou HanFa, and Lu HaoJie, have made huge efforts in the development of materials and technology for identification and quantification of protein glycosylation. Among these, the N-linked glycoproteome has been extensively characterized. To enrich these N-linked peptides, magnetic nanoparticles and a monolithic capillary column-based glycoproteomic reactor have been developed by Zou HanFa's and Lu HaoJie's groups, respectively, for enriching N-linked glycopeptides with high efficiency [43–46]. In order to analyze the N-glycosylated proteome more efficiently, integrated rapid and sensitive platforms have also been developed by several labs for analysis of serum or tissues samples [47–49]. Besides, to investigate the abundance and the biological function of core-fucosylated glycoproteins, a multiple reaction monitoring-based quantitative proteomics approach has been developed for site-specific quantification of these glycoproteins by Zhao Yan et al. from BPRC [50].

With the exception of glycosylated protein substrates and their specific modified sites, glycans may also serve as markers. Therefore, glycans from modified protein substrates have also been extensively characterized. Zhang WanJun from Qian XiaoHong's lab [49] developed a highly efficient method for glycan enrichment. Zou HanFa's lab [52] enriched N-linked glycans using OMC. Yang Peng-Yuan's lab [53] developed a method to incorporate ^{18}O into glycans with endoglycosidase to relatively quantify glycans on modified substrates. These studies greatly enhanced glycoproteomics and biomarker studies in China.

2.2.3 Technology development for other protein PTMs

Ubiquitination is also one of the common PTMs of lysine in protein substrates, and regulates the biological function, or even determines the fate of the modified substrates. Protein ubiquitination is thought to be specifically determined by ubiquitin ligases (E3s). However, the substrates of most E3s have not been extensively explored. Gao YouHe and his colleagues [54] developed a proteomics tool to identify E3-specific substrates by affinity purification with protein interaction domains recognizable by E3. They also verified the ubiquitination features using an *in vitro* assay. This strategy can potentially be adapted to any E3 with a protein interaction domain(s), and may serve as a powerful tool for the comprehensive identification of E3 substrates on a proteomic scale. With the exception of substrate specificity, the specific ubiquitin chains also play regulatory roles in the biological functions of their modified protein substrates. Using an MS-based approach, Li Lin and his colleagues [55] demonstrated that Smurf1-mediated Lys29-linked polyubiquitination of axin negatively regulates Wnt/β-catenin signaling without proteolytic activity.

Acetylation is another common and biologically important PTM. However, the study of the acetylome primarily depends on the development of an affinity approach for acetylated peptides with low affinity in the cell. Zhao Shi-Min and his colleagues [56] successfully developed a pan-acetyllysine antibody and a site-specific antibody to Lys288-acetylated argininosuccinate lyase using chemically acetylated ovalbumin and synthetic acetylated peptide. Using these antibodies in a proteomics approach, the group purified and characterized the acetylome after subcellular fractionation of human liver tissue, with reduced contamination with highly abundant acetylated histones.

2.3 Materials and technology development in protein identification

Although the improvement in MS performance has made it possible to achieve deep coverage of proteomes, the high performance of the MS platform also benefits from the optimized use of chromatographic systems. Multidimensional chromatography coupled with tandem MS is useful for large-scale proteomic analysis. In order to optimize this method, Ding Chen et al. [57] at BPRC developed a fast workflow for identification and quantification of proteomes that could provide a strategy that effectively balances efficiency and depth in proteome coverage. They used a dual RP-HPLC-MS/MS approach with a short gradient to achieve deep coverage of the proteome. Using this approach, they identified about 8000 gene products in human cells in just 12 h of MS running time. This "Fast-seq" approach can eliminate the desalting step and may be compatible with any label-free or label-based absolute or relative quantification. Also in phosphoproteome research, Zou HanFa and his colleagues [36] established a new RP-RP liquid chromatography approach for multidimensional separation of phosphopeptides. By applying the new strategy, about 8000 phosphorylation sites in human liver samples were identified by maintaining protein false discovery rate lower than 1%.

2.4 Quantitative proteomics and its application in biological research

In the past years, several novel quantitative labeling methods have been reported. For example, Song ChunXia et al.

created a new strategy to improve quantification accuracy and throughput for phosphoproteome analysis. They developed a pseudo-triplex stable isotope dimethyl labeling approach coupled with an online RP-strong cation exchange-RP multidimensional separation system. The throughput and accuracy of the quantitative analysis were improved significantly by this strategy combined with the relative standard deviation criterion [38]. Finally, over 1800 phosphopeptides corresponding to 1918 unique phosphorylation sites were reliably quantified in a 42-hour online multidimensional analysis. Nie AiYing et al. [58] developed a novel quantitative strategy for *in vivo* terminal amino acid labeling. With the utilization of the heavy amino acids $^{13}C_6$-arginine and $^{13}C_6$-lysine, and specific endoproteinases Lys-N and Arg-C, they generated some labeled isobaric peptides that are indistinguishable in the MS scan, but exhibit multiple MS/MS reporter b and y ion pairs in a full mass range. The novel strategy generated an abundance of reliable quantitative information, high sensitivity, and good dynamic range of nearly two orders of magnitude. Another interesting study was implemented by Liu SiQi and his colleagues. They applied another isobaric tag, deuterium isobaric amine reactive tag (DiART), for quantifying the proteome of *Thermoanaerobacter tengcongensis*. They found that DiART performed well in ion intensity and quantification accuracy compared with iTRAQ [59]. Also, Zou HanFa's group developed the isobaric peptide terminal labeling strategy using formaldehyde labeled with different isotopes for site-selective dimethyl labeling of N-terminal peptides. They obtained more than 98% selectivity and 99% labeling efficiency for complex samples [60]. In addition, Ge Feng and his colleagues applied quantitative proteomics in identification of microRNA (miR) targets. By knocking down endogenous miR-21 in U266 myeloma cells, they employed a stable isotope labeling of amino acids in cell culture (SILAC)-based quantitative proteomic strategy to systematically identify potential targets of miR-21. Because miRNAs may regulate many of their targets at the translational level without affecting mRNA abundance, proteomic methods are best suited for revealing the full spectrum of miRNA targets [61,62]. In addition, Zeng Rong and her colleagues [63] firstly used the SILAC mouse strategy to screen diagnostic markers in the serum and urine of immunoglobulin A nephropathy patients, and then found related novel candidates, such as complement C3, albumin, vitamin D binding protein (VDBP), ApoA1, and insulin-like growth factor binding protein 7 (IGFBP7), which might provide potential disease-related biomarkers for evaluation of treatment Besides, Zhang LinJun's group [64] also discovered that the expression of protein disulfide-isomerase associated protein 3 (PDIA3) was decreased in non-parenchymal cells of their rat liver fibrosis models by 2D-gel combined LC-MS/MS.

2.5 Bioinformatics development

With the development of experimental technology, a large amount of proteomics data was generated rapidly, which makes data analysis challenging. Algorithms and applications focusing on data quality control, *de novo* sequencing, genome re-annotation and PTM identification were developed to provide high fidelity proteomics data, from which novel proteins/PTMs could be discovered. Besides, other applications and databases were developed for the purpose of gene/protein annotation and biomarker discovery.

2.5.1 Bioinformatics tools for protein sequencing

In a data-dependent approach, the first step in interpreting mass spectra is determining the monoisotopic peak of a precursor, which is critical for accurate database matching. By combining theoretical isotopic distribution and similarity of elution profiles, Yuan ZuoFei et al. [65] developed pParse, which recalls many more precursor ions with high accuracy, resulting in a higher peptide identification rate. At BPRC, Li Ning et al. developed a quality control tool, PepDistiller, to improve the sensitivity and accuracy of peptide identification for MASCOT search. Besides, multithreading technology was applied to accelerate the data processing [66]. In order to improve the efficiency of large-scale proteomics studies, BuildSummary was developed to increase sensitivity during database matching [67]. Furthermore, Li YiXue and his colleagues [68] at East China University of Science and Technology developed a pattern feature-matching algorithm, which separates correct identifications and random matches clearly, with the highest values for sensitivity (82%), precision (97%), and F1-measure (89%) of protein identification. Although these studies successfully improved the power of protein identification at each individual step, further studies are necessary to integrate all of these tools for highly efficient and accurate searching in protein identification.

De novo sequencing is another protein identification tool for extracting peptide sequences directly from tandem MS data without a database search. Recent advances in MS technology have made it possible to dissect the spectrum with complementary fragmentation information, high resolution, and high mass accuracy with the support of higher energy collisional dissociation (HCD) and electron transfer dissociation (ETD). He Simin's and Dong Mengqiu's lab pioneered this direction in the field. They developed the search engine pNovo+ for *de novo* peptide sequencing with a greater than 70% success rate using HCD and ETD alone or complementary HCD and ETD tandem mass spectra [69–71]. More interestingly, they used pNovo+ to help identify a member of the serine protease inhibitor (Serpin) family, As_SRP-1, which is secreted from spermatids during spermiogenesis in the nematode *Ascaris suum*. This information will be valuable in deciphering molecular mechanisms of sperm maturation in Nematodes [72].

Collaborating in carrying out MS experiments and bioinformatics analysis, Dong's and He's groups [73] developed pLink software for cross-link identification. They demonstrated that pLink reliably estimates false discovery rate in cross-link identification and is a robust tool for protein structure exploration.

2.5.2 Bioinformatics tools for genome re-annotation

In the post-genomic era, proteomics analysis does not only consist of identifying and quantitating proteins using a data-dependent approach with the support of the annotated genome, but also consist of calibrating gene structure annotation. Li YiXue's and Xie Lu's groups at the Shanghai Center for Bioinformation Technology tested this possibility with high mass accuracy MS data from mouse samples. They also developed an iGepros server to provide biologists with integrated large-scale genes and proteins annotation [74,75]. At the same time, He Fuchu, Yang PengYuan, Li SiQi, and He QingYu represented CNHUPO and chaired chromosomes 1, 8, and 20 for the Chromosome-centric Human Proteome Project (C-HPP), respectively, and the project was guided by HE FuChu, the executive committee member of C-HPP. The aim of this project is to confirm all of the coding genes in the genome using a proteomics approach and re-annotate the human genome. The portion implemented by CNHUPO occupied about one fifth of this international project, showing the determination and confidence of Chinese proteomists to go from following the lead of international colleagues to taking the lead in an international collaboration [76–79].

2.5.3 Identification of protein PTMs

Identification of proteins and their modifications through LC-MS/MS is an important task in the proteomics field. However, because of the complexity of tandem mass spectra and the data-dependent searching algorithm with a limited number of known modifications, the majority of the modified spectra cannot be identified. He SiMin and his colleagues developed the statistical algorithm DeltAMT for fast detection of abundant protein modifications by comparing the simultaneously presented, modified and unmodified versions of a peptide from the same batch of tandem mass spectra with high-accuracy precursor masses. This highly efficient, accurate, and sensitive method provides a useful tool for yielding deep insights into the data and significantly increasing the spectral identification rates [80].

With the exception of global notification of PTMs, significant effort has also been focused on the data analysis for multiple important PTMs. For example, for thephosphoproteome data analysis, Jiang XinNing et al. [81] found that different classes of mass spectra were significantly different in the pattern and quality of phosphopeptide spectra, with the result that phosphopeptide identifications from these classes were highly complementary. By classifying mass spectra into different groups and setting special criteria for phosphopeptides identified from different classes of mass spectra, higher sensitivity and more comprehensive coverage were achieved.

2.5.4 Bioinformatics tools for protein-protein interaction

Protein-protein interaction studies greatly facilitate exploration of biological functions in life science through the context of the protein environment. Among them, self-interacting proteins (for which two or more copies may interact) play important roles in understanding the cellular functions of the proteins and the evolution of protein interaction networks. Liu ZhongYang et al. developed SLIPPER (SeLf-Interacting Protein PrEdictoR), a classification system that integrates several functional annotation and network topology features to predict self-interacting proteins and fill in the gaps between bioinformatics prediction and high-throughput experimental demonstration of self-interacting proteins [82]. By utilizing 5-fold cross-validation and an independent test, the authors convinced the good performance of this model. Furthermore, they provided the user-friendly web service SLIPPER for this prediction model.

2.5.5 Database development for proteomics studies

In addition to the development of data processing software, in the past years, multiple labs in China developed several databases to support large-scale proteomics analyses. Lu Xie's group developed a database of differentially expressed proteins in human cancers (dbDEPC) by curating proteomics data of 20 types of cancers. The database provides a resource for information on protein-level expression changes and exploring protein profile differences among different cancers [83]. Yang PengYuan et al. [84] at Fudan University released nuclear proteome profile of C57BL/6J mouse liver, in which a total of 748 proteins were identified. They also provided a promising strategy to enrich low abundance proteins especially transcription factors. Shao Chen et al. at Peking Union Medical College developed Urinary Protein Biomarker for biomarker screening. In this database, they collected over 500 records of human biomarkers from existing urinary protein biomarker studies in the published literature after manual curation [85]. These powerful tools can serve as references for both proteomics and cancer researchers.

With regard to PTMs, Xue Yu and his colleagues developed several databases for the different types of PTM analysis. For example, they manually collected 26 E1s; 105 E2s; 1003 E3s; and 148 deubiquitinating enzymes for ubiquitin and ubiquitin-like conjugation systems in the scientific literature [86]. Also, they classified E3s into a hierarchical structure with five levels, i.e., class, group, subgroup, family, and single E3. Finally, they created the comprehensive ubiquitin and ubiquitin-like conjugation database (UUCD), which contains 56949 enzymes for ubiquitin and ubiquitin-like conjugation across 70 eukaryotic species. In addi-

tion to UUCD, they created the MiCroKit 3.0 database, which provides detailed information for 1489 proteins from seven different model organisms [87]. They also created the compendium of protein lysine acetylation (CPLA) 1.0 of lysine-acetylated substrates and their acetylation sites [88].

3 Perspectives

With consistent policy and foundation support over the past three years, large improvements have been made by the Chinese proteomics community. The development of proteomics technologies, including the sample preparation and enrichment methods for special PTMs or low-abundance proteins, improved digestion approaches and ionization methods, the optimized LC-MS/MS system, and standardized data processing software, will all make it possible to generate comprehensive descriptions of the multiple-organ proteome of human beings and other model organisms through proteome deep sequencing and global quantitation, which makes proteomics a powerful tool to study the mechanisms underlying physiological/pathological processes of life.

We appreciate the great funding support from the Chinese central and local governments, especially from the Ministry of Science and Technology of China, National Natural Science Foundation of China, and the Beijing City. This work was funded by the Chinese National Basic Research Program grants (2011CB910600, 2013CB911201) to Xu Ping and Li Ning respectively and the grants of National Natural Science Foundation of China (31070673, 31170780) to Xu Ping. We also appreciate the dedication of all of the members of CNHUPO and Chinese proteomists, especially the authors of the references. Owing to time and space limitations, we may have omitted some representative works from our CNHUPO society.

1 Gao X, Zhang X, Zheng J, He F. Proteomics in China: Ready for prime time. Sci China Life Sci, 2010, 53: 22–33
2 He F. At a glance: Proteomics in China. Sci China Life Sci, 2011, 54: 1–2
3 He F. Lifeomics leads the age of grand discoveries. Sci China Life Sci, 2013, 56: 201–212
4 Leng F. Opportunity and challenge: Ten years of proteomics in China. Sci China Life Sci, 2012, 55: 837–839
5 He F. Human liver proteome project: Plan, progress, and perspectives. Mol Cell Proteomics, 2005, 4: 1841–1848
6 Zhang X, Guo Y, Song Y, Sun W, Yu C, Zhao X, Wang H, Jiang H, Li Y, Qian X, Jiang Y, He F. Proteomic analysis of individual variation in normal livers of human beings using difference gel electrophoresis. Proteomics, 2006, 6: 5260–5268
7 Sun A, Jiang Y, Wang X, Liu Q, Zhong F, He Q, Guan W, Li H, Sun Y, Shi L, Yu H, Yang D, Xu Y, Song Y, Tong W, Li D, Lin C, Hao Y, Geng C, Yun D, Zhang X, Yuan X, Chen P, Zhu Y, Li Y, Liang S, Zhao X, Liu S, He F. Liverbase: A comprehensive view of human liver biology. J Proteome Res, 2010, 9: 50–58
8 Consortium CHLPP. First insight into the human liver proteome from proteomesky-liverhu 1.0, a publicly available database. J Proteome Res, 2010, 9: 79–94
9 Zhao S, Xu W, Jiang W, Yu W, Lin Y, Zhang T, Yao J, Zhou L, Zeng Y, Li H, Li Y, Shi J, An W, Hancock SM, He F, Qin L, Chin J, Yang P, Chen X, Lei Q, Xiong Y, Guan KL. Regulation of cellular metabolism by protein lysine acetylation. Science, 2010, 327: 1000–1004
10 Wang Q, Zhang Y, Yang C, Xiong H, Lin Y, Yao J, Li H, Xie L, Zhao W, Yao Y, Ning ZB, Zeng R, Xiong Y, Guan KL, Zhao S, Zhao GP. Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. Science, 2010, 327: 1004–1007
11 Song C, Ye M, Liu Z, Cheng H, Jiang X, Han G, Songyang Z, Tan Y, Wang H, Ren J, Xue Y, Zou H. Systematic analysis of protein phosphorylation networks from phosphoproteomic data. Mol Cell Proteomics, 2012, 11: 1070–1083
12 Wang J, Huo K, Ma L, Tang L, Li D, Huang X, Yuan Y, Li C, Wang W, Guan W, Chen H, Jin C, Wei J, Zhang W, Yang Y, Liu Q, Zhou Y, Zhang C, Wu Z, Xu W, Zhang Y, Liu T, Yu D, Zhang Y, Chen L, Zhu D, Zhong X, Kang L, Gan X, Yu X, Ma Q, Yan J, Zhou L, Liu Z, Zhu Y, Zhou T, He F, Yang X. Toward an understanding of the protein interaction network of the human liver. Mol Syst Biol, 2011, 7: 536
13 Liu S, Chen H, Lu X, Deng C, Zhang X, Yang P. Facile synthesis of copper(II) immobilized on magnetic mesoporous silica microspheres for selective enrichment of peptides for mass spectrometry analysis. Angew Chem Int Ed Engl, 2010, 49: 7557–7561
14 Wan H, Qin H, Xiong Z, Zhang W, Zou H. Facile synthesis of yolk-shell magnetic mesoporous carbon microspheres for efficient enrichment of low abundance peptides. Nanoscale, 2013, 5: 10936–10944
15 Ding C, Chan DW, Liu W, Liu M, Li D, Song L, Li C, Jin J, Malovannaya A, Jung SY, Zhen B, Wang Y, Qin J. Proteome-wide profiling of activated transcription factors with a concatenated tandem array of transcription factor response elements. Proc Natl Acad Sci USA, 2013, 110: 6771–6776
16 Liu Q, Ding C, Liu W, Song L, Liu M, Qi L, Fu T, Malovannaya A, Wang Y, Qin J, Zhen B. In-depth proteomic characterization of endogenous nuclear receptors in mouse liver. Mol Cell Proteomics, 2013, 12: 473–484
17 Qin H, Gao P, Wang F, Zhao L, Zhu J, Wang A, Zhang T, Wu R, Zou H. Highly efficient extraction of serum peptides by ordered mesoporous carbon. Angew Chem Int Ed Engl, 2011, 50: 12218–12221
18 Jian G, Liu Y, He X, Chen L, Zhang Y. Click chemistry: A new facile and efficient strategy for the preparation of Fe_3O_4 nanoparticles covalently functionalized with IDA-Cu and their application in the depletion of abundant protein in blood samples. Nanoscale, 2012, 4: 6336–6342
19 Zhang M, He X, Chen L, Zhang Y. Preparation and characterization of iminodiacetic acid-functionalized magnetic nanoparticles and its selective removal of bovine hemoglobin. Nanotechnology, 2011, 22: 065705
20 Liu Y, Xue Y, Ji J, Chen X, Kong J, Yang P, Hubert HH, Liu B. Gold nanoparticle assembly microfluidic reactor for efficient on-line proteolysis. Mol Cell Proteomics, 2007, 1428–1436
21 Zhang X, Zhu S, Xiong Y, Deng C. Development of a MALDI-TOF MS strategy for the high-throughput analysis of biomarkers: On-target aptamer immobilization and laser-accelerated proteolysis. Angew Chem Int Ed Engl, 2013, 52: 6055–6058
22 Yao G, Deng C, Zhang X, Yang P. Efficient tryptic proteolysis accelerated by laser radiation for peptide mapping in proteome analysis. Angew Chem Int Ed Engl, 2010, 49: 8185–8189
23 Bian Y, Ye M, Song C, Cheng K, Wang C, Wei X, Zhu J, Chen R, Wang F, Zou H. Improve the coverage for the analysis of phosphoproteome of hela cells by a tandem digestion approach. J Proteome Res, 2012, 11: 2828–2837
24 Qin W, Song Z, Fan C, Zhang W, Cai Y, Zhang Y, Qian X. Trypsin immobilization on hairy polymer chains hybrid magnetic nanoparticles for ultra fast, highly efficient proteome digestion, facile ^{18}O labeling and absolute protein quantification. Anal Chem, 2012, 84: 3138–3144
25 Chen H, Deng C, Zhang X. Synthesis of Fe_3O_4@SiO_2@PMMA

core-shell-shell magnetic microspheres for highly efficient enrichment of peptides and proteins for maldi-tof ms analysis. Angew Chem Int Ed Engl, 2010, 49: 607–611
26 Sun L, Ma J, Qiao X, Liang Y, Zhu G, Shan Y, Liang Z, Zhang L, Zhang Y. Integrated device for online sample buffer exchange, protein enrichment, and digestion. Anal Chem, 2010, 82: 2574–2579
27 Pan Y, Ye M, Zhao L, Cheng K, Dong M, Song C, Qin H, Wang F, Zou H. N-terminal labeling of peptides by trypsin-catalyzed ligation for quantitative proteomics. Angew Chem Int Ed Engl, 2013, 52: 9205–9209
28 Dong M, Wu M, Wang F, Qin H, Han G, Dong J, Wu R, Ye M, Liu Z, Zou H. Coupling strong anion-exchange monolithic capillary with MALDI-TOF MS for sensitive detection of phosphopeptides in protein digest. Anal Chem, 2010, 82: 2907–2915
29 Nie S, Dai J, Ning ZB, Cao XJ, Sheng QH, Zeng R. Comprehensive profiling of phosphopeptides based on anion exchange followed by flow-through enrichment with titanium dioxide (AFET). J Proteome Res, 2010, 9: 4585–4594
30 Zhang Y, Chen C, Qin H, Wu R, Zou H. The synthesis of Ti-hexagonal mesoporous silica for selective capture of phosphopeptides. Chem Commun, 2010, 46: 2271–2273
31 Yan Y, Zheng Z, Deng C, Li Y, Zhang X, Yang P. Hydrophilic polydopamine-coated graphene for metal ion immobilization as a novel immobilized metal ion affinity chromatography platform for phosphoproteome analysis. Anal Chem, 2013, 85: 8483–8487
32 Xu B, Zhou L, Wang F, Qin H, Zhu J, Zou H. Selective capture of phosphopeptides by hierarchical Ti-aluminophosphate-5 molecular sieves. Chem Commun, 2012, 48: 1802–1804
33 Zhang L, Zhao Q, Liang Z, Yang K, Sun L, Zhang L, Zhang Y. Synthesis of adenosine functionalized metal immobilized magnetic nanoparticles for highly selective and sensitive enrichment of phosphopeptides. Chem Commun, 2012, 48: 6274–6276
34 Li QR, Ning ZB, Tang JS, Nie S, Zeng R. Effect of peptide-to-TiO_2 beads ratio on phosphopeptide enrichment selectivity. J Proteome Res, 2009, 8: 5375-5381
35 Zhao X, Wang Q, Wang S, Zou X, An M, Zhang X, Ji J. Citric acid-assisted two-step enrichment with TiO_2 enhances the separation of multi- and monophosphorylated peptides and increases phosphoprotein profiling. J Proteome Res, 2013, 12: 2467–2476
36 Song C, Ye M, Han G, Jiang X, Wang F, Yu Z, Chen R, Zou H. Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. Anal Chem, 2010, 82: 53–56
37 Qin H, Wang F, Wang P, Zhao L, Zhu J, Yang Q, Wu R, Ye M, Zou H. Phosphoric acid functionalized mesoporous organo-silica (epo) as the adsorbent for *in situ* enrichment and isotope labeling of endogenous phosphopeptides. Chem Commun, 2012, 48: 961–963
38 Song C, Wang F, Ye M, Cheng K, Chen R, Zhu J, Tan Y, Wang H, Figeys D, Zou H. Improvement of the quantification accuracy and throughput for phosphoproteome analysis by a pseudo triplex stable isotope dimethyl labeling approach. Anal Chem, 2011, 83: 7755–7762
39 Jiang X, Ye M, Han G, Dong X, Zou H. Classification filtering strategy to improve the coverage and sensitivity of phosphoproteome analysis. Anal Chem, 2010, 82: 6168–6175
40 Zhu J, Wang F, Cheng K, Song C, Qin H, Hu L, Figeys D, Ye M, Zou H. Analysis of human serum phosphopeptidome by a focused database searching strategy. J Proteomics, 2013, 78: 389–397
41 Wang Z, Ding G, Geistlinger L, Li H, Liu L, Zeng R, Tateno Y, Li Y. Evolution of protein phosphorylation for distinct functional modules in vertebrate genomes. Mol Biol and Evol, 2011, 28: 1131–1140
42 Ren J, Jiang C, Gao X, Liu Z, Yuan Z, Jin C, Wen L, Zhang Z, Xue Y, Yao X. PhosSNP for systematic analysis of genetic polymorphisms that influence protein phosphorylation. Mol Cell Proteomics, 2010, 9: 623–634
43 Xiong Z, Qin H, Wan H, Huang G, Zhang Z, Dong J, Zhang L, Zhang W, Zou H. Layer-by-layer assembly of multilayer polysaccharide coated magnetic nanoparticles for the selective enrichment of glycopeptides. Chem Commun, 2013, 49: 9284–9286
44 Zhang Y, Kuang M, Zhang L, Yang P, Lu H. An accessible protocol for solid-phase extraction of N-linked glycopeptides through reductive amination by amine-functionalized magnetic nanoparticles. Anal Chem, 2013, 85: 5535–5541
45 Liu J, Wang F, Lin H, Zhu J, Bian Y, Cheng K, Zou H. Monolithic capillary column based glycoproteomic reactor for high-sensitive analysis of N-glycoproteome. Anal Chem, 2013, 85: 2847–2852
46 Xiong Z, Zhao L, Wang F, Zhu J, Qin H, Wu R, Zhang W, Zou H. Synthesis of branched peg brushes hybrid hydrophilic magnetic nanoparticles for the selective enrichment of N-linked glycopeptides. Chem Commun, 2012, 48: 8138–8140
47 Ma C, Zhao X, Han H, Tong W, Zhang Q, Qin P, Chang C, Peng B, Ying W, Qian X. N-linked glycoproteome profiling of human serum using tandem enrichment and multiple fraction concatenation. Electrophoresis, 2013, 34: 2440–2450
48 Zhu J, Wang F, Chen R, Cheng K, Xu B, Guo Z, Liang X, Ye M, Zou H. Centrifugation assisted microreactor enables facile integration of trypsin digestion, hydrophilic interaction chromatography enrichment, and on-column deglycosylation for rapid and sensitive N-glycoproteome analysis. Anal Chem, 2012, 84: 5146–5153
49 Qu Y, Xia S, Yuan H, Wu Q, Li M, Zou L, Zhang L, Liang Z, Zhang Y. Integrated sample pretreatment system for n-linked glycosylation site profiling with combination of hydrophilic interaction chromatography and PNGase F immobilized enzymatic reactor via a strong cation exchange precolumn. Anal Chem, 2011, 83: 7457–7463
50 Zhao Y, Jia W, Wang J, Ying W, Zhang Y, Qian X. Fragmentation and site-specific quantification of core fucosylated glycoprotein by multiple reaction monitoring-mass spectrometry. Anal Chem, 2011, 83: 8802–8809
51 Zhang W, Han H, Bai H, Tong W, Zhang Y, Ying W, Qin W, Qian X. A highly efficient and visualized method for glycan enrichment by self-assembling pyrene derivative functionalized free graphene oxide. Anal Chem, 2013, 85: 2703–2709
52 Qin H, Zhao L, Li R, Wu R, Zou H. Size-selective enrichment of N-linked glycans using highly ordered mesoporous carbon material and detection by MALDI-TOF MS. Anal Chem, 2011, 83: 7721–7728
53 Zhang W, Wang H, Tang H, Yang P. Endoglycosidase-mediated incorporation of ^{18}O into glycans for relative glycan quantitation. Anal Chem, 2011, 83: 4975–4981
54 Guo Z, Song E, Ma S, Wang X, Gao S, Shao C, Hu S, Jia L, Tian R, Xu T, Gao Y. Proteomics strategy to identify substrates of LNX, a PDZ domain-containing E3 ubiquitin ligase. J Proteome Res, 2012, 11: 4847–4862
55 Fei C, Li Z, Li C, Chen Y, Chen Z, He X, Mao L, Wang X, Zeng R, Li L. Smurf1-mediated lys29-linked nonproteolytic polyubiquitination of axin negatively regulates wnt/beta-catenin signaling. Mol Cell Biol, 2013, 33: 4095–4105
56 Guan KL, Yu W, Lin Y, Xiong Y, Zhao S. Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides. Nat Protoc, 2010, 5: 1583–1595
57 Ding C, Jiang J, Wei J, Liu W, Zhang W, Liu M, Fu T, Lu T, Song L, Ying W, Chang C, Zhang Y, Ma J, Wei L, Malovannaya A, Jia L, Zhen B, Wang Y, He F, Qian X, Qin J. A fast workflow for identification and quantification of proteomes. Mol Cell Proteomics, 2013, 12: 2370–2380
58 Nie AY, Zhang L, Yan GQ, Yao J, Zhang Y, Lu HJ, Yang PY, He FC. *In vivo* termini amino acid labeling for quantitative proteomics. Anal Chem, 2011, 83: 6026–6033
59 Chen Z, Wang Q, Lin L, Tang Q, Edwards JL, Li S, Liu S. Comparative evaluation of two isobaric labeling tags, diart and itraq. Anal Chem, 2012, 84: 2908–2915
60 Qin H, Wang F, Zhang Y, Hu Z, Song C, Wu R, Ye M, Zou H. Isobaric cross-sequence labeling of peptides by using site-selective N-terminus dimethylation. Chem Commun, 2012, 48: 6265–6267

61 Xiong Q, Zhong Q, Zhang J, Yang M, Li C, Zheng P, Bi LJ, Ge F. Identification of novel miR-21 target proteins in multiple myeloma cells by quantitative proteomics. J Proteome Res, 2012, 11: 2078–2090

62 Li C, Xiong Q, Zhang J, Ge F, Bi LJ. Quantitative proteomic strategies for the identification of microRNA targets. Expert Rev Proteomics, 2012, 9: 549–559

63 Zhao S, Li R, Cai X, Chen W, Li Q, Xing T, Zhu W, Chen YE, Zeng R, Deng Y. The application of SILAC mouse in human body fluid proteomics analysis reveals protein patterns associated with IgA nephropathy. Evid Based Complement Alternat Med, 2013, 2013: 275390

64 Zhao Q, Feng Y, Jia X, Yin L, Zheng Y, Ouyang D, Zhou H, Zhang L. Proteome analysis of hepatic non-parenchymal cells of immune liver fibrosis rats. Sci China Life Sci, 2014, 57: 303–314

65 Yuan ZF, Liu C, Wang HP, Sun RX, Fu Y, Zhang JF, Wang LH, Chi H, Li Y, Xiu LY, Wang WP, He SM. Pparse: A method for accurate determination of monoisotopic peaks in high-resolution mass spectra. Proteomics, 2012, 12: 226–235

66 Li N, Wu S, Zhang C, Chang C, Zhang J, Ma J, Li L, Qian X, Xu P, Zhu Y, He F. Pepdistiller: A quality control tool to improve the sensitivity and accuracy of peptide identifications in shotgun proteomics. Proteomics, 2012, 12: 1720–1725

67 Sheng Q, Dai J, Wu Y, Tang H, Zeng R. Buildsummary: Using a group-based approach to improve the sensitivity of peptide/protein identification in shotgun proteomics. J Proteome Res, 2012, 11: 1494–1502

68 Li Y, Hao P, Zhang S, Li Y. Feature-matching pattern-based support vector machines for robust peptide mass fingerprinting. Mol Cell Proteomics, 2011, 10: M110 005785

69 Chi H, Chen H, He K, Wu L, Yang B, Sun RX, Liu J, Zeng WF, Song CQ, He SM, Dong MQ. pNovo+: *De novo* peptide sequencing using complementary HCD and ETD tandem mass spectra. J Proteome Res, 2013, 12: 615–625

70 Chi H, Sun RX, Yang B, Song CQ, Wang LH, Liu C, Fu Y, Yuan ZF, Wang HP, He SM, Dong MQ. pNovo: *De novo* peptide sequencing and identification using HCD spectra. J Proteome Res, 2010, 9: 2713–2724

71 Sun RX, Dong MQ, Song CQ, Chi H, Yang B, Xiu LY, Tao L, Jing ZY, Liu C, Wang LH, Fu Y, He SM. Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra. J Proteome Res, 2010, 9: 6354–6367

72 Zhao Y, Sun W, Zhang P, Chi H, Zhang MJ, Song CQ, Ma X, Shang Y, Wang B, Hu Y, Hao Z, Huhmer AF, Meng F, L'Hernault S W, He SM, Dong MQ, Miao L. Nematode sperm maturation triggered by protease involves sperm-secreted serine protease inhibitor (Serpin). Proc Natl Acad Sci USA, 2012, 109: 1542–1547

73 Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, Li S, Chi H, Li YX, Chen HF, Luo SK, Ding YH, Wang LH, Hao Z, Xiu LY, Chen S, Ye K, He SM, Dong MQ. Identification of cross-linked peptides from complex samples. Nat Methods, 2012, 9: 904–906

74 Xing XB, Li QR, Sun H, Fu X, Zhan F, Huang X, Li J, Chen CL, Shyr Y, Zeng R, Li YX, Xie L. The discovery of novel protein-coding features in mouse genome based on mass spectrometry data. Genomics, 2011, 98: 343–351

75 Zheng G, Wang H, Wei C, Li Y. Igepros: An integrated gene and protein annotation server for biological nature exploration. BMC Bioinformatics, 2011, 12(Suppl 14): S6

76 Wu S, Li N, Ma J, Shen H, Jiang D, Chang C, Zhang C, Li L, Zhang H, Jiang J, Xu Z, Ping L, Chen T, Zhang W, Zhang T, Xing X, Yi T, Li Y, Fan F, Li X, Zhong F, Wang Q, Zhang Y, Wen B, Yan G, Lin L, Yao J, Lin Z, Wu F, Xie L, Yu H, Liu M, Lu H, Mu H, Li D, Zhu W, Zhen B, Qian X, Qin J, Liu S, Yang P, Zhu Y, Xu P, He F. First proteomic exploration of protein-encoding genes on chromosome 1 in human liver, stomach, and colon. J Proteome Res, 2013, 12: 67–80

77 Zhang Y, Yan G, Zhai L, Xu S, Shen H, Yao J, Wu F, Xie L, Tang H, Yu H, Liu M, Yang P, Xu P, Zhang C, Li L, Chang C, Li N, Wu S, Zhu Y, Wang Q, Wen B, Lin L, Wang Y, Zheng G, Zhou L, Lu H, Liu S, He F, Zhong F. Proteome atlas of human chromosome 8 and its multiple 8p deficiencies in tumorigenesis of the stomach, colon, and liver. J Proteome Res, 2013, 12: 81−88

78 Wang Q, Wen B, Yan G, Wei J, Xie L, Xu S, Jiang D, Wang T, Lin L, Zi J, Zhang J, Zhou R, Zhao H, Ren Z, Qu N, Lou X, Sun H, Du C, Chen C, Zhang S, Tan F, Xian Y, Gao Z, He M, Chen L, Zhao X, Xu P, Zhu Y, Yin X, Shen H, Zhang Y, Jiang J, Zhang C, Li L, Chang C, Ma J, Yan G, Yao J, Lu H, Ying W, Zhong F, He QY, Liu S. Qualitative and quantitative expression status of the human chromosome 20 genes in cancer tissues and the representative cell lines. J Proteome Res, 2013, 12: 151−161

79 Guo F, Wang D, Liu Z, Lu L, Zhang W, Sun H, Zhang H, Ma J, Wu S, Li N, Jiang Y, Zhu W, Qin J, Xu P, Li D, He F. Caper: A chromosome-assembled human proteome browser. J Proteome Res, 2013, 12: 179−186

80 Fu Y, Xiu LY, Jia W, Ye D, Sun RX, Qian XH, He SM. Deltamt: A statistical algorithm for fast detection of protein modifications from L-MS/MS data. Mol Cell Proteomics, 2011, 10: M110.000455

81 Jiang X, Ye M, Han G, Dong X, Zai H. Classification filtering strategy to improve the coverage and sensitivity of phosphoproteome analysis. Anal Chem, 2010, 82: 6168–6175

82 Liu Z, Guo F, Zhang J, Wang J, Lu L, Li D, He F. Proteome-wide prediction of self-interacting proteins based on multiple properties. Mol Cell Proteomics, 2013, 12: 1689–1700

83 He Y, Zhang M, Ju Y, Yu Z, Lv D, Sun H, Yuan W, He F, Zhang J, Li H, Li J, Wang-Sattler R, Li Y, Zhang G, Xie L. Dbdepc 2.0: Updated database of differentially expressed proteins in human cancers. Nucleic Acids Res, 2012, 40: 964–971

84 Zhang Y, Fang C, Bao H. Nuclear proteome profile of C57Bl/6J mouse liver. Sci China Life Sci, 2013, 56: 513–523

85 Shao C, Li M, Li X, Wei L, Zhu L, Yang F, Jia L, Mu Y, Wang J, Guo Z, Zhang D, Yin J, Wang Z, Sun W, Zhang Z, Gao Y. A tool for biomarker discovery in the urinary proteome: A manually curated human and animal urine protein biomarker database. Mol Cell Proteomics, 2011, 10: M111.010975

86 Gao T, Liu Z, Wang Y, Cheng H, Yang Q, Guo A, Ren J, Xue Y. UUCD: A family-based database of ubiquitin and ubiquitin-like conjugation. Nucleic Acids Res, 2013, 41: 445–451

87 Ren J, Liu Z, Gao X, Jin C, Ye M, Zou H, Wen L, Zhang Z, Xue Y, Yao X. Microkit 3.0: An integrated database of midbody, centrosome and kinetochore. Nucleic Acids Res, 2010, 38: 155–160

88 Liu Z, Cao J, Gao X, Zhou Y, Wen L, Yang X, Yao X, Ren J, Xue Y. Cpla 1.0: An integrated database of protein lysine acetylation. Nucleic Acids Res, 2011, 39: 1029–1034