Chapter 2
Proteomics Defines Protein Interaction Network of Signaling Pathways

Shitao Li

Abstract Protein interactions play fundamental roles in signaling transduction. Analysis of protein–protein interaction (PPI) has contributed numerous insights to the understanding of the regulation of signal pathways. Different approaches have been used to discover PPI and characterize protein complexes. In addition to conventional PPI methods, such as yeast two-hybrid (YTH), affinity purification coupled with mass spectrometry (AP-MS) is emerging as an important and popular tool to unravel protein complex and elucidate protein function through the interaction partners. With the AP-MS method, protein complexes are prepared first by affinity purification directly from cell lysates, followed by characterization of their components by mass spectrometry. In contrast to most PPI methods, AP-MS reflects PPI under near physiological conditions in the relevant organism and cell type. AP-MS is also able to probe dynamic PPI dependent on protein posttranslational modifications, which is common for signal transduction. AP-MS mapping protein interaction network of various signal pathways has dramatically increased in recent years. Here, I'll present the strategies toward obtaining an interactome map of signal pathway and the methodology, detailed protocols, and perspectives of AP-MS.

Keywords Protein interactions • Signaling transduction • Mass spectrometry • Affinity purification • Interaction network • Dynamic

S. Li, Ph.D. (✉)
Department of Microbiology and Immunobiology,
Harvard Medical School, Boston, MA 02115, USA
e-mail: lishitao@hotmail.com

X. Wang (ed.), *Bioinformatics of Human Proteomics*, Translational Bioinformatics 3,
DOI 10.1007/978-94-007-5811-7_2,

2.1 Introduction

Protein interaction plays essential role in cell structure and function. In a simplified diagram of a signaling pathway, upon interaction of a ligand, the receptor alters its conformation, such as dimerization, phosphorylation, and ubiquitination, leading to recruitment of intracellular molecules and subsequent activation of downstream signal cascades. Each level of the signaling cascades requires protein interaction to work as a well-assembled, multifunctional protein complex essential for signal transduction. The functionality of proteins relies on their ability to interact with one another, whereas pathogenic conditions can reflect the perturbations of these protein interactions.

Numerous protein–protein interaction (PPI) methods have been developed, but only a few of them are used for large-scale PPI detection, including yeast two-hybrid (YTH), protein fragment complementation assay (PCA), luciferase-mediated interactome (LUMIER), mammalian protein–protein interaction trap (MAPPIT), protein array, and affinity purification coupled with tandem mass spectrometry (AP-MS). The YTH system is the first assay for analysis of large-scale protein–protein interactions and widely accepted method (Fields and Song 1989). In YTH system, interested gene (bait, X) is fused to the DNA-binding (DB) domain of a transcription factor such as Gal4 (DB-X), while the interacting protein (prey, Y) is fused to an activation domain (AD) such as Gal4-AD (AD-Y). Physical interaction between X and Y brings AD and DB together, which reconstitutes the transcription factor and subsequently activates the downstream reporter genes (Fields and Song 1989). Like the YTH, PCA requires that bait and prey are each fused with incomplete fragments of a third protein, which acts as a reporter. Interaction between bait and prey proteins brings the fragments of reporter protein in close enough proximity to allow them to form a functional reporter protein (Rossi et al. 1997). When fluorescent proteins are reconstituted, the PCA is called bimolecular fluorescence complementation assay (Kerppola 2009). LUMIER is basically a co-immunoprecipitation assay, in which bait is linked to an epitope for purification and prey protein is fused to renilla or firefly luciferase for detection (Barrios-Rodiles et al. 2005). In the MAPPIT, bait and prey proteins are linked to signaling deficient cytokine receptor chimeras. Interaction of bait and prey restores JAK–STAT cascade after the receptor has been stimulated with ligand, which leads to STAT3-dependent reporter gene activation (Eyckerman et al. 2001). Protein microarray is a microscopic array glass slide on which interested proteins have been affixed at separate locations in an ordered manner using a variety of available chemical linkers (MacBeath 2002). Protein microarrays are typically high-density arrays that are used to identify novel proteins or protein–protein interactions. Antibody microarrays are the most common analytical microarray.

AP-MS is biochemical purification of protein complexes followed by characterization of their components by mass spectrometry. However, unlike the methods discussed above, AP-MS is not designed for one-to-one protein interaction (i.e., binary interaction). Instead, AP-MS detects multi-protein complexes. As with

AP-MS, gene of interests is tagged with desirable epitope for affinity purification. Various tags have been developed, such as FLAG tag, HA tag, glutathione S-transferase (GST) tags, the calmodulin-binding peptide, the streptavidin-binding peptide, or the in vivo biotinylation of the target tagged peptide using coexpression of the BirA ligase (Waugh 2005). With affinity tag, protein complexes are enriched first by affinity purification. One early developed AP-MS is to use the tandem affinity purification (TAP) tag (Puig et al. 2001). The original TAP tag is composed of a protein A tag and a calmodulin-binding peptide for two sequential enrichment purifications. In the first purification step, the protein complex is isolated from the cell lysate using immunoglobulin gamma (IgG) resin with high protein A affinity. After protein complex is cleaved from the protein A tag with TEV protease, the eluate undergoes second purification on an immobilized calmodulin column.

To date, AP-MS has been performed in combination with other techniques, such as biochemical fractionation and chemical cross-linking, for characterization of protein complex. Combining biochemical fractionations, like size fractionation, with AP-MS can provide a more precise characterization of multi-protein complexes according to the factions. For example, a combination of TAP purification with standard gel filtration has allowed for a better characterization of RNA polymerase II complex (Mueller and Jaehning 2002). Cross-linker is used for detecting weak interactions, such as membrane complex, which may be interrupted by detergents in lysis buffer. A combination of TAP with in vivo cross-linking with formaldehyde was used to identify novel proteasome interactors (Tagwerker et al. 2006). AP-MS can also be combined with quantitative proteomics approaches, such as SILAC and ICAT, to better understand the dynamics of protein complex assembly. Stable isotope labeling by amino acids in cell culture (SILAC) is an approach for in vivo incorporation of a label into proteins for mass spectrometry (MS)-based quantitative proteomics (Ong et al. 2002). Isotope-coded affinity tags (ICAT) are complementary to SILAC and measure dynamic changes in complexes isolated from tissues or organisms that cannot be metabolically labeled (Gygi et al. 1999). Both entail labeling the samples with isotope labels that allow the mass spectrometer to distinguish between identical proteins in separate samples. Differentially labeled samples are combined and analyzed together, and the differences in the peak intensities of the isotope pairs accurately reflect difference in the abundance of the corresponding proteins.

Given the fundamental importance of protein interactions, systematically mapping protein–protein interaction (PPI) in various species has dramatically increased in recent years. Using high-throughput YTH, proteome-wide physical interaction maps have been generated for several organisms: *Saccharomyces cerevisiae* (Fromont-Racine et al. 1997; Uetz et al. 2000; Ito et al. 2001), *Caenorhabditis elegans* (Walhout et al. 2000; Reboul et al. 2003; Li et al. 2004), *Drosophila melanogaster* (Giot et al. 2003; Guruharsha et al. 2011), and human (Guruharsha et al. 2011; Rual et al. 2005). Virus–host protein interactomes were also explored, such as severe acute respiratory syndrome (SARS)-coronavirus (Pfefferle et al. 2011), Kaposi sarcoma herpesvirus (KSHV), and Varicella zoster virus (VZV) (Uetz

et al. 2006; Rozen et al. 2008). In addition to global mapping, protein interaction networks of several important signal pathways, such as MAPK (Bandyopadhyay et al. 2010), TGFβ (Tewari et al. 2004), SMAD (Colland et al. 2004), and PI3K-mTOR (Pilot-Storck et al. 2010), have been investigated.

In addition to YTH, AP-MS is another widely used PPI tool to map protein interactomes. Due to many advantages that will be discussed later, AP-MS mapping protein interaction network of various signal pathways has dramatically increased in recent years. Global-wide interactomes have been established in *Escherichia coli* (Hu et al. 2009) and *Mycoplasma pneumonia* (Kuhner et al. 2009), *Saccharomyces cerevisiae* (Krogan et al. 2006; Gavin et al. 2006; Ho et al. 2002), *Drosophila melanogaster* (Guruharsha et al. 2011), and HIV–host interactome (Jager et al. 2012). In vertebrate, this approach has so far been used to define proteomic subspaces or specific signal pathways: antiviral innate immunity pathway (Li et al. 2011), autophagy pathway (Behrends et al. 2010), deubiquitinase interactome (Sowa et al. 2009), endoplasmic reticulum-associated protein degradation network (ERAD) (Christianson et al. 2012), TNF pathway (Bouwmeester et al. 2004), proteasome interaction network (Guerrero et al. 2008), and disease-related protein network (Ewing et al. 2007).

Systematic identification of protein interactions within an organism will facilitate systems-level studies of biological processes. Current binary PPI networks are mainly generated by high-throughput yeast two-hybrid. Due to the small overlap of these maps, it has been assumed that these maps are of low quality containing many false positives (Parrish et al. 2006). Recent efforts to map interactions using AP-MS illustrate the promise to measure specific protein interactions in vivo (instead of in yeast) and provide a more powerful tool to model the in vivo interactome. First, I discuss the advantages of AP-MS versus YTH, and then focus the details of the methodology, applications, and perspectives of AP-MS.

2.2 AP-MS Versus YTH

Despite the wide acceptance of YTH system for protein–protein interaction analysis and discovery, high-throughput YTH for protein interaction network bears several major limitations: (1) Reporter analysis method indirectly reflects protein–protein interaction which usually leads to high false positives. For example, proteins with transcriptional activity can lead to autoactivation of the reporter genes. (2) Some heterologous protein expressions are incompatible or toxic to yeast, i.e., membrane proteins which are unlikely to be appropriately assayed as a fusion with a reconstituted transcription factor in YTH. (3) YTH cannot reflect the endogenous protein interactions in the relevant organism. (4) Lots of signaling pathways in vertebrates do not exist in yeast. Thus, interactions triggered by posttranslational modifications do not occur in yeast, resulting in many intrinsic false negatives. (5) The coverage of prey library usually is not completed. In addition, in high-throughput YTH, the bait expression is not monitored. Heterologous full-length protein expression,

Table 2.1 Comparison between YTH and AP-MS

	Cell type	Interaction detection	Interaction type	Interaction level	Interaction status	Cost
YTH	Yeast	Indirect	Binary	Overexpression	Static	Relatively cheap
AP-MS	Relevant species	Direct	Multi-complexes	Endogenous	Static/dynamic	Expensive

especially high-molecular-weight protein, expects to have low expression level in yeast.

Although both YTH and AP-MS detect protein–protein interaction, they have several distinct differences (Table 2.1). AP-MS couples affinity purification with mass spectrometry and requires more labor works and sophisticated equipments. Basically, baits can be expressed in any cell line, which investigator is interested in. After antibiotic selection, bait expression levels are monitored in stable cell lines by western blot, and cell line expressing low bait protein level (close to endogenous level) is usually chosen for following affinity purification. Since the bait expression is close to the counterpart endogenous protein level, we expect the purified complex reflects the endogenous protein interactions under physiological conditions. AP-MS also can be used to detect dynamic protein interactions dependent on protein posttranslational modification by signal stimulation. Unlike YTH detecting one-to-one interaction (aka binary interaction), AP-MS analyzes the entire bait complex and provides all prey information in one run. However, the purified complex represents a mix of direct and indirect binding partners since the nature of the interactions identified in AP-MS data cannot be determined to be either direct or indirect. Last, protein abundance and specificity in different cell lines also limits the detection of protein complex. For example, MIB1 and MIB2 have comparable affinity with TBK1, but we did not detect MIB2 in TBK1 complex in 293T cells by AP-MS. Using real-time PCR, we found MIB1 predominantly expressed in 293T cell line (Li et al. 2011). Taken all together, AP-MS overcomes the limitations of YTH discussed above except several disadvantages over YTH: high cost, indirect interaction, and cell type specificity.

2.3 Methodology of AP-MS

The pipeline of AP-MS from gene construction to interaction network mapping is shown in Fig. 2.1 (Li et al. 2011). In brief, interested gene is tagged with desirable epitopes such as FLAG, GST, His, and biotin. Depending on the purification strategy, one or two tags (usually tandem tags) are adopted. These vectors should carry one antibiotic resistance gene for mammalian cell stable line selection.

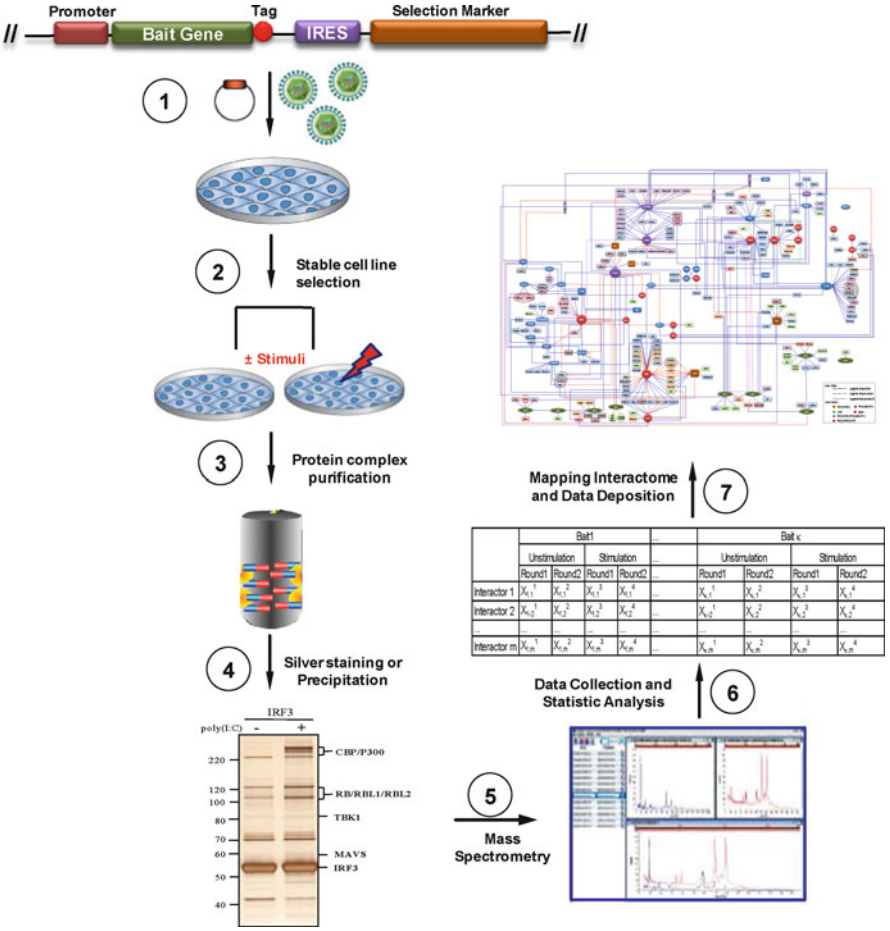

Fig. 2.1 Schematic illustration of the experimental pipeline from gene construction, stable cell line selection, and protein complex affinity purification and identification to data analysis and interactome mapping

After transfection or infection into the desirable mammalian cell line, cells are selected by designated antibiotics to obtain stably and close to endogenous protein expression. Protein complexes are precipitated from lysates of bulk cells by using various immobilized matrixes, such as resin conjugated with antibody. Protein complexes are then eluated from the matrixes after several washing steps to remove nonspecific interactors. Protein complex is either separated on gel following silver staining or precipitated. Sliced gel bands or solution samples are analyzed by mass spectrometry. After data collection and statistical analysis, protein interaction network is generated and ready for validation and further function analysis.

2.3.1 Vector

To purify protein complex closing to physiological level, cell line stably expressing tagged bait is a prerequisite. Therefore, antibiotic resistance gene should be included in the vector for stable cell line selection. Genes of interest also needs to be tagged in-frame with an epitope (at either the N or C terminus), which is used to affinity purify the tagged protein (aka bait) along with its interacting partners (aka prey). Any affinity tag can be used for AP-MS in theory, and most successful tags developed to date are FLAG, HA, S-tag, and tandem affinity purification (TAP) tag. Each purification tag has advantages and disadvantages, and the appropriate technique should be selected depending on the goals of the experiment. For example, a single FLAG or HA epitope only adds 8–11 amino acids (Li et al. 2011), while the TAP tag adds a >20-kDa tag (Krogan et al. 2006) which may cause more nonspecific binding. Because tag may interfere with protein expression or interaction, both N-terminal and C-terminal fusion could be tested for optimal AP-MS. For example, membrane protein may need to put the tag on the C-terminal or after signal peptide on the N-terminus. Furthermore, two kinds of purification methods (single and tandem purification) are used for AP-MS, which requires bait fused with single or double epitopes, respectively.

2.3.2 Purification

Depending on the number of tags on the vector, there are one-step and two-step purification methods for specific protein complex, cell line, or organism. Originally developed for yeast, the first TAP tag consists of calmodulin-binding peptide (CBP), followed by tobacco etch virus protease (TEV protease) cleavage site and protein A with high affinity to immunoglobulin gamma (IgG). Protein complex is first purified from the cell lysate on an IgG affinity resin and cleaved from the protein A tag with TEV protease. The eluate is then enriched in a second affinity purification step on an immobilized calmodulin column. Several variants of TAP with different combinations of tags, such as FLAG-HA double tags, are developed.

Usually, one-step purifications on average preserve weaker or more transient protein–protein interactions in the price of a higher number of nonspecific binding proteins. Conversely, the tandem procedure tends to yield cleaner results, but weak interactions can be lost. FLAG and HA double tags are most commonly applied for tandem purification of protein complexes. We compared the effect of tandem tag versus single tag purification on the yield of total prey and HCIP by examining four protein complexes purified by single purification with FLAG versus a two-step purification with FLAG followed by HA (Li and Dorf 2013). MS analysis revealed that the number of total interactors was dramatically reduced in all protein complexes (TBK1, NAP1, IRF3, and SINTBAD) isolated by TAP purification. However, the ratio of HCIP to total prey did not increase. Consistently, more HCIP were detected

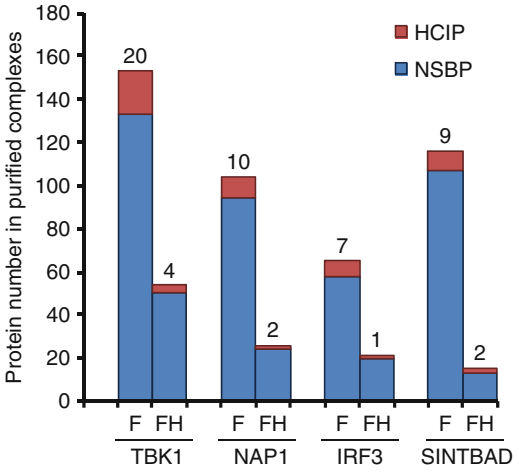

Fig. 2.2 Comparison of one-step and tandem purifications. Protein numbers identified in *TBK1*, *NAP1*, *IRF3*, and *SINTBAD* complexes are depicted. *F* stands for FLAG affinity purification, and *FH* indicates FLAG and HA tandem purification

by single-step affinity purification (Fig. 2.2). In brief, tandem purification reduces the NSBP at the price of HCIP loss. Due to on average more than 90% of proteins as nonspecific binding protein in one-step purification, researchers prefer to tandem affinity purification to get a cleaner background if they only study on a few protein complexes. However, if the study is to map the protein interaction network of a specific signaling pathway, NSBP from one-step purification can be excluded by statistical analysis of the whole database.

2.3.3 Mass Spectrometric Protein Identification

In most proteomics experiments, the purified proteins are separated by one-dimensional SDS-PAGE and stained with a mass spectrometry-compatible dye such as silver, SYPRO ruby, or Coomassie. SDS-PAGE separation removes unwanted contaminants such as buffer components from the protein sample, and the sample complexity is decreased by separating the proteins according to molecular weight. Moreover, it also can be used to compare bands distribution with and without stimulation. In some cases, like IRF3 complexes shown in Fig. 2.1, unique bands are only found in the bait complex with stimulation, indicating these interacting proteins are dependent on ligand stimulation.

Individual protein bands of interest are excised, or the entire lane is cut into approximately 1-mm^3 pieces. Gel pieces were then subjected to an in-gel trypsin digestion procedure to produce peptides for mass spectrometry analysis. But the extraction efficiency of peptides from a gel is low and dependent on the primary structure of the peptide. As an alternative approach to in-gel digestion, protein mixtures can be digested in solution without prior separation (Behrends et al. 2010).

Because buffer components, such as detergents, interfere with the mass spectrometry ionization process, protein samples need to be precipitated with trichloroacetic acid (TCA), washed, and redissolved in a digestion buffer. The main advantages of solution digestion are the reduction of the time and a higher recovery of peptides compared to in-gel digestion. However, bear in mind that some proteins like membrane proteins are resistant to be redissolved.

The peptide mixture can be directly introduced into the mass spectrometer or separated by HPLC before mass spectrometric analysis (LC-MS). The two primary mass spectrometry methods developed for identification of proteins are electrospray ionization (ESI) (Fenn et al. 1989) and matrix-assisted laser desorption/ionization (MALDI) (Hillenkamp et al. 1991). Electrospray ionization mass spectrometry is a desorption ionization method. A sample solution is sprayed from a small tube into a strong electric field in the presence of a flow of warm nitrogen to assist desolvation. The droplets formed evaporate in a region maintained at a vacuum of several torr causing the charge to increase on the droplets. The multiply charged ions then enter the analyzer. The most obvious feature of an ESI spectrum is that the ions carry multiple charges, which reduces their mass-to-charge ratio compared to a singly charged species. This advantage allows mass spectra to be obtained for large molecules. A major disadvantage is that this technique cannot analyze mixtures very well. The other most used technique, MALDI, is a two-step process. First, desorption is triggered by a UV laser beam. Matrix material heavily absorbs UV laser light, leading to the ablation of upper layer (~micron) of the matrix material. A hot plume produced during the ablation contains many species: neutral and ionized matrix molecules, protonated and deprotonated matrix molecules, matrix clusters, and nanodroplets. The second step is ionization (more accurately protonation or deprotonation). In the most common instrumental designs, ESI and MALDI are performed with mass spectrometers capable of tandem mass spectrometry (MS/MS) experiments. Ion traps, quadrupole time-of-flight instruments (Q-TOF), Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers (FTMS), and the Orbitrap are the most common types of instrumentation now used in high-end protein analysis.

2.3.4 Quantification and Dynamics

Most protein interactomes only represent as static entities, which however only poorly captures the dynamics of complex composition. There has been increasing efforts to detect dynamic views of interactomes using various modified AP-MS. Systematic methods to map dynamic changes include semi-quantification based on total spectral counts or ion intensities of precursor peptide (MS1) or fragment ions (MS2) and use of isotopic labeling approaches to obtain more accurate relative quantification. Relative quantification methods such as the stable isotope labeling by amino acids in cell culture (SILAC) detect differences in protein abundance among samples using nonradioactive isotopic labeling. Although relative quantitation

is more costly and time-consuming, and less sensitive to experimental bias than label-free quantitation, it entails labeling the samples with stable isotope labels that allow the mass spectrometer to distinguish between identical proteins in separate samples. Differentially labeled samples are combined and analyzed together, and the differences in the peak intensities of the isotope pairs accurately reflect difference in the abundance of the corresponding proteins. Thus, relative quantitation may discover the dynamic interactions by comparing the change of identical protein abundances from same bait cells with and without extracellular stimulation. Absolute quantitation of proteins is also developed by using isotopic peptides entails spiking known concentrations of synthetic, heavy isotopologues of target peptides into an experimental sample (Mirgorodskaya et al. 2012). However, the cost of absolute quantitation is too high and not realistic for large-scale interactome mapping.

As quantitative methods become more robust, there will be increasing demand for detection of dynamic protein interaction upon extracellular stimulation. For example, we revealed that ~20% protein interactions are dependent on ligand stimulation, such as viral dsRNA mimics poly(dI:dC), in the *h*uman *i*nnate *i*mmunity *i*nteractome for type *I i*nterferon (HI5) (Li et al. 2011). Another example in insulin pathway, Glatter et al. defined the interaction network of insulin receptor/target of rapamycin pathway in Drosophila (Glatter et al. 2011). They found that 22% of the detected interactions were regulated by insulin. In addition to the quantitative power of mass spectrometry, it is also crucial to establish a stable cell line sensitive to stimulations. When overexpressed in cells, bait protein may not respond to stimuli as sensitive as the corresponding endogenous protein.

2.3.5 Data Collection and Analysis

In most cases, the raw data files are first processed by the software controlling the respective mass spectrometry instrument. The generated data sets are then searched against a protein database using search engines such as MASCOT (Hirosawa et al. 1993) or SEQUEST (MacCoss et al. 2002). A valid approach for validation of the chosen parameters is to search the obtained data sets against a decoy protein database. The data also need to be further filtered by setting specific thresholds such as a minimum peptide length or a specific number of peptides to consider a protein identification. Mass spectrometry has some intrinsic problems, such as the common problem of carryovers between mass spectrometry runs. To circumvent the carryover problem in mass spectrometry, we usually analyze the repeated sample in different batch. The carryovers in two independent AP-MS of the same bait will not be possible to show up twice. The record of each batch of MS runs will also help to discriminate the carryovers.

In addition to mass spectrometry, affinity purification also has its own inherent false positives and false negatives, which is critical general limitation encountered in the interpretation of the AP-MS due to lack of binary interaction information. False positives are nonspecific binding proteins and contaminants found in purified bait complex. Several types of false positives are present in typical affinity purified protein samples. The most

common ones are from researchers' hands when they perform purification and handle samples. These contaminants usually are keratin proteins and easy to remove from the dataset. There are also other various kinds of nonspecific binding proteins: (1) proteins binding to affinity matrices, like STK38 and PRMT5; (2) proteins bind to affinity tag, like KIF11 binding to FLAG tag; (3) abundant proteins (e.g., actin, tubulin); (4) proteins prefer binding to specific domain, like ribosomal proteins binding to baits with nucleic acid-binding domain; (5) and heat-shock proteins for protein folding. Therefore, it is important to use cell line stably expressing baits at near physiological levels to avoid NSBPs, as transient overexpression may probably result in protein aggregation and improper intracellular localization. To discriminate NSBP from the protein complex, repetition of AP-MS is mandatory. In our experiences, NSBPs are dramatically different in two independent AP-MS of the same bait. Proper controls including cells expressing GFP with the same epitope will be also useful to exclude NSBPs. Last, large database with the same affinity tag and the same cell line background from high-throughput study will be a good resource for identification of NSBPs and HCIPs. If a protein is often isolated with many unrelated bait proteins, it is easily recognized through analysis of the high-throughput data. However, systematic large-scale experiment does not allow for the subjective and individual evaluation of their results, which means the removal of potential contaminating proteins cannot be based on judging individual purifications. Therefore, statistic tools for analysis of database are required to filter out nonspecific proteins and yield high-confidence interacting proteins.

For statistical analysis of AP-MS data, three main parameters are protein abundance, uniqueness (the frequency of observed protein in database), and reproducibility. Total spectral counts (TSC) have gained acceptance as a practical, label-free, semiquantitative measure of protein abundance in proteomics study. Several computational tools have been developed for the processing of AP-MS data, like CompPASS (Sowa et al. 2009), SAINT (Breitkreutz et al. 2010), and MiST (Jager et al. 2012).

We designed a simplified method for analysis of AP-MS data, combining three main parameters: protein abundance, uniqueness (the frequency of observed protein in the database), and reproducibility. Total spectral counts (TSC) have gained acceptance as a practical, label-free, semiquantitative measure of protein abundance for proteomics studies. We adopted the *z*-score statistic to compare protein abundance because *z*-score calculates the probability of TSC occurring within a normal distribution. However, *z*-score does not reflect reproducibility. In our protocol, each protein complex is tested in 4 MS runs, so reproducibility can be readily factored into the analysis. *z*-Score also does not analyze information about prey occurrence (i.e., prey uniqueness). To explore the likelihood that an interaction is specific, we set a value of prey occurrence at <5%. We now propose a simple 3-stage scoring system to identify HCIP. This algorithm combines *z*-score plus prey occurrence and reproducibility (ZSPORE) (Li and Dorf 2013). In the ZSPORE scoring system, each interaction must pass all three criteria to merit classification as HCIP. The flowchart of ZSPORE is shown as in Fig. 2.3, and a detailed description is provided in Sect. 2.4.6. Taken together, the ZSPORE method combines three parameters (*z*-score based on TSC, prey occurrence, and reproducibility) and is a simple, efficient, and robust way to analyze AP-MS data.

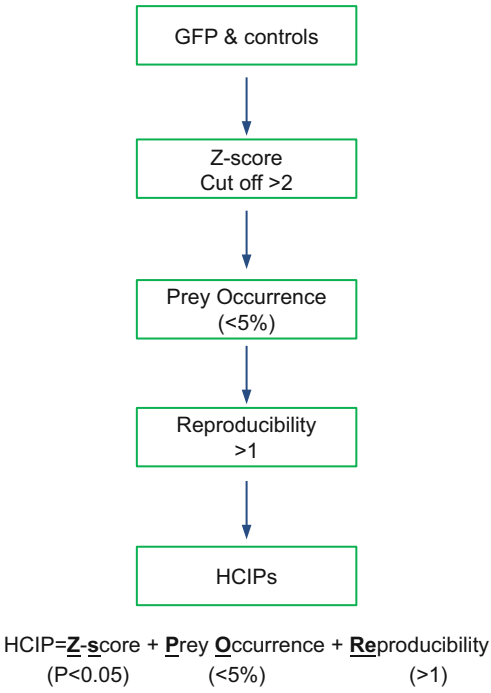

Fig. 2.3 Flowchart of ZSPORE analysis

As with any large screening database, AP-MS also has false negatives, like lacking many known protein–protein interactions documented previously. There are several reasons why a known interaction fail to be found in AP-MS. First, statistical analysis tool may filter out the known interaction as a nonspecific binding. Second, the nature and location of the tag might interfere bait protein function and disrupt its interactions. Third, to parallel comparison, all AP-MS experiments are performed in a same single condition. The generic conditions of affinity purification may be too harsh to preserve some protein interactions, such as the buffer for membrane proteins should be different from other ones. Fourth, the known protein interaction depends on different stimulation. Some proteins may be involved in several pathways and have different interactors in response to the relevant stimulation. Last, the absence of detection is often due to the protein expression level in the specific cell type, especially when the cells have relative low abundances of the protein.

2.3.6 Network Mapping and Analysis

To visualize the protein interaction network formed by HCIPs and baits, graphic representation of two protein interactions basically consists of drawing two circles (nodes) linked by a line (edge). All interactions are combined to generate a map of

the protein interaction network or interactome. A common protein interactome displays a few highly connected nodes forming hubs or subnetwork, while most nodes have a few edges. Several software are developed for graphic mapping protein interaction network. Cytoscape is the most used open-source software platform, which can be used to visualize complex networks derived from AP-MS data. Cytoscape is available for free download at http://www.cytoscape.org. A lot of plug-ins are also available for various kinds of problem domains, including bioinformatics, social network analysis, and semantic web (Smoot et al. 2011). For comprehensive and dynamic visualization of the network, various kinds of attributes can be applied to the node and the edge by representation of different color and line thickness.

In addition, the functional classifications of HCIPs can be analyzed by a few online programs. For example, HCIP list can be uploaded to PANTHER (Thomas et al. 2003) or DAVID (da Huang et al. 2009) via a web interface. These programs group these proteins by protein domains, molecular functions, biological processes, and signal pathways. The functional classifications may help discover common threads underlying the proteins of interest. Another approach is to obtain clues from known protein interactions to discover regulation mechanisms. Several protein–protein interaction databases are available for online search, repository, and free download, such as BioGRID, STRING, IntAct, and MINT. The BioGRID database is an online protein interaction repository with data compiled through comprehensive curation efforts. The latest version searches 31,739 publications for 510,188 raw protein and genetic interactions from major model organism species (Stark et al. 2011). The STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations. STRING quantitatively integrates interaction data from these sources for a large number of organisms and transfers information between these organisms where applicable (Szklarczyk et al. 2011). The IntAct database provides a freely available, open-source database system and analysis tools for molecular interaction data (Kerrien et al. 2012). All interactions are derived from literature curation or direct user submissions and are freely available. The MINT database focuses on experimentally verified protein–protein interactions mined from the scientific literature by expert curators (Licata et al. 2012).

AP-MS raw data also can be deposited in the Tranche repository (Smith et al. 2011), which is a distributed file system into which any sort of proteomics data may be uploaded. The data then are distributed on the internet and downloaded by anyone who has access to the hash key identifiers for the data, which may be kept private or publicly released. In summary, all these free online programs are useful and convenient research tools for mapping, analysis, and repository of AP-MS data.

2.4 Protocols

AP-MS has applied for mapping of protein interactome of various cellular signaling pathways in mammalian cells. Our lab has established an efficient AP-MS pipeline for defining protein interaction network and successfully applied in several

pathways including *h*uman *i*nnate *i*mmunity *i*nteractome for type *I* *i*nterferon (HI5) (Li et al. 2011), *mi*RNA pathway *i*nteractome (Mii), and influenza–host (iHost) protein interaction network (Li and Dorf, unpublished data). Detailed pipeline of our AP-MS is provided in this section, and how this applies on different pathways in mammalian cells will be discussed.

2.4.1 Bait Selection and cDNA Cloning

Genes known to regulate the studied signaling pathway are usually selected as primary baits. Baits cover from extracellular signals like ligands to cognate receptors on cell membrane and to signaling intermediates, kinases, and transcription factors involved in these signaling pathways and their family members. After analysis of primary bait AP-MS, some new and important HCIPs with primary baits are also chosen to be as secondary baits. Secondary baits will validate the association with primary baits but also expand the protein interaction network, provide new insights into this signaling pathway, and cross talk with other pathways.

Bait cDNAs can be tagged with various epitopes, such as FLAG or HA epitope. As we discussed earlier, commercially available anti-FLAG beads have much higher affinity than anti-HA beads. We use two mammalian expression vectors, pCMV-3Tag8 (Stratagene) and viral expression vector, pLPCX (Clontech), for transfection and infection, respectively. Vector pCMV-3Tag8 harbors a hygromycin resistance gene, while pLPCX confers cells' resistance to puromycin.

2.4.2 Establishment of Stable Cell Line and Cell Stimulation

Transfection and transduction are two common DNA delivery methods into mammalian cells. For cell lines easy to be transfected like HEK293 cells, bait constructs are directly transfected into cells. For cell lines with low transfection efficiency, such as THP-1 cell line, bait gene needs to be first packaged into retroviral virion. The following infection will allow bait gene to integrate into cell genome DNA and subsequent expression in cells. Two days after transfection and infection, cells are treated with puromycin or hygromycin for 14 days. Single colonies are picked and expanded in 6-well plates. Protein expression levels in each colony are determined by immunoblotting. Colony with protein expression close to endogenous level is picked up for AP-MS.

Most protein interactomes are descriptions of homeostasis of a specific signaling pathway, such as DUB network (Sowa et al. 2009), autophagy interaction network (Behrends et al. 2010), and ERAD interactome (Christianson et al. 2012). However, many protein interactions depend on protein posttranslational modifications induced by different stimuli. For example, we found that about 20% interactions were ligand dependent in HI5 protein interaction network (Li et al. 2011). We also noticed many

new interactions between influenza virus protein and human host after viral infection (Li and Dorf, unpublished data). Therefore, in our pipeline for AP-MS, each stable cell line is divided into two groups, and cells are treated with ligand specific for the signaling pathway or infected with virus for studying virus–host interactome.

2.4.3 Complex Purification

Each group of cells is cultured in four or five 15-cm^2 culture dishes (about 5×10^7 cells) to scale up for affinity purification. Cells are lysed in 10 ml TAP buffer (50 mM Tris HCl [pH 7.5], 10 mM $MgCl_2$, 100 mM NaCl, 0.5% Nonidet P40, 10% glycerol, phosphatase inhibitors, and protease inhibitors). After shaking on ice for 30 min, cell lysates were centrifuged for 30 min at 15,000 rpm. Supernatants are collected and precleared with 50 µl of protein A/G resin. After shaking for 1 h at 4°C, resin is removed by centrifugation. Cell lysates are added to 20 µl anti-FLAG M2 resin (Sigma) and incubated on a shaker for 12 h. Then the anti-FLAG resin is 3× washed (15 min/time) with 10 ml TAP buffer. After removing the wash buffer, the resin is transferred to a spin column (Sigma) and incubated with 40 µl 3× FLAG peptide (Sigma) for 1 h at 4°C in a shaker. Eluates are collected by centrifugation and stored at −80°C.

2.4.4 Silver Staining

Purified complexes are loaded on 4–15% NuPAGE gels (Invitrogen) and run about 1 cm^2 distance for 8 min at 200 V. Gels were stained using the SilverQuest Staining Kit (Invitrogen). Each entire stained lane was excised and rinsed twice with 50% acetonitrile.

2.4.5 Mass Spectrometry

The Taplin Biological Mass Spectrometry Facility (Harvard Medical School) performs MS analysis for our samples. Excised gel bands were cut into approximately 1-mm^3 pieces. Gel pieces are then subjected to a modified in-gel trypsin digestion procedure. Gel pieces were washed and dehydrated with acetonitrile for 10 min followed by removal of acetonitrile. Pieces were then completely dried in a speed-vac. Gel pieces were rehydrated with 50 mM ammonium bicarbonate solution containing 12.5 ng/µl modified sequencing grade trypsin (Promega, Madison, WI) at 4°C. After 45 min, the excess trypsin solution was removed and replaced with 50 mM ammonium bicarbonate solution to just cover the gel pieces. Peptides were later

extracted by removing the ammonium bicarbonate solution, followed by one wash with a solution containing 50% acetonitrile and 1% formic acid. The extracts were then dried in a speed-vac (~1 h) and stored at 4°C until analysis.

On the day of analysis, the samples were reconstituted in 5–10 μl of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nanoscale reverse-phase HPLC capillary column was created by packing 5 μm C18 spherical silica beads into a fused silica capillary (100-μm inner diameter *x* ~12-cm length) with a flame-drawn tip. After equilibrating the column, each sample was loaded via a FAMOS auto sampler (LC Packings, San Francisco, CA) onto the column. A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid).

As peptides eluted, they were subjected to electrospray ionization and then entered into an LTQ Velos ion trap mass spectrometer (Thermo Fisher, San Jose, CA). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Dynamic exclusion was enabled such that ions were excluded from reanalysis for 30 s. Peptide sequences (and hence protein identity) were determined by matching protein databases with the acquired fragmentation pattern by the software program SEQUEST (Thermo Fisher, San Jose, CA). The human IPI database (ver. 3.6) was used for searching. Precursor mass tolerance was set to ±2.0 Da, and MS/MS tolerance was set to 1.0 Da. A reversed-sequence database was used to set the false discovery rate at 1%. Filtering was performed using the SEQUEST primary score, Xcorr, and delta-Corr. Spectral matches were further manually examined, and multiple identified peptides (>1) per protein were required.

2.4.6 Statistical Analysis of Mass Spectrometry Data

As with many screening methods, unfiltered AP-MS data contain many nonspecific binding proteins due to some intrinsic characteristics, such as nonspecific binding to bead or tag, protein aggregation, and carryover during MS runs. We now describe a simple efficient statistic method, *z*-score plus prey occurrence and reproducibility (ZSPORE) scoring system, for identification of HCIP. Using this pipeline, we achieve a higher efficiency of AP-MS and better identification of high-confidence interacting proteins. The methods and criteria used to remove nonspecific binding proteins and identify high-confidence interacting proteins include:

(a) GFP and controls. AP-MS of GFP-FLAG and various controls, such as non-FLAG IgG conjugated resin for AP-MS, were used to identify nonspecific binding proteins in the database.
(b) *z*-Score. A *z*-score (aka a standard score) indicates how many standard deviations an element is from the mean. To calculate *z*-score, mass spectrometry data were transformed into a "stats table," where the columns are total spectral counts (TSC) from 4 MS runs, the rows are bait-associated proteins (Table 2.2). Then we calculated *z*-score of each $X_{i,j}$ (i prey interacts with j bait) based on

Table 2.2 Statistical analysis table

	Bait 1				Bait κ			
	Unstimulation		Stimulation		Unstimulation		Stimulation	
	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2
Interactor 1	$X^1_{1,1}$	$X^2_{1,1}$	$X^1_{2,1}$	$X^2_{2,1}$	$X^1_{2\kappa-1,1}$	$X^2_{2\kappa-1,1}$	$X^1_{2\kappa,1}$	$X^2_{2\kappa,1}$
Interactor 2	$X^1_{1,2}$	$X^2_{1,2}$	$X^{s1}_{2,2}$	$X^{s2}_{2,2}$	$X^1_{2\kappa-1,2}$	$X^2_{2\kappa-1,2}$	$X^1_{2\kappa,2}$	$X^2_{2\kappa,2}$
Interactor m	$X^1_{1,m}$	$X^2_{1,m}$	$X^1_{2,m}$	$X^2_{2,m}$	$X^1_{2\kappa-1,m}$	$X^2_{2\kappa-1,m}$	$X^1_{2\kappa,m}$	$X^2_{2\kappa,m}$

the maximum total spectral counts (TSC) of 4 MS runs. For HI5 database analysis, we set the cutoff of z-score as 2.

$$z = \frac{(X - \mu)}{\sigma}$$

z is the z-score, X is the value of the element, μ is the population mean, and σ is the standard deviation.

(c) Prey occurrence. We considered any prey associated with a single bait as an HCIP while preys associated with all baits as NSBP. Generally, we set the bar of prey occurrence as <5%, which means one specific prey interacts less than 5% of total baits in the entire database. In HI5, we showed that preys that interact with less than 5 baits represented statistically significant interactions in HI5 dataset. So the threshold for prey occurrence in HI5 is set as 4. Due to known high interconnectivity among selected baits, bait-to-bait interactions were considered as HCIP.

(d) Reproducibility. Each prey must appear in at least 2 out of 4 MS runs.

(e) Batch reproducibility. To account for possible variations in the list of background contaminants observed in our dataset that were not identified by other statistical approaches, we intentionally sequenced each duplicate purified complex in different experiments. Any protein that did not appear in different purifications was considered an NSBP and manually removed from HCIP list.

2.4.7 Construction of Protein Interaction Map and Bioinformatics Analysis

After statistical analysis of dataset, all pairwise interactions are collected and analyzed by Cytoscape. Several important attributes, such as z-score and TSC, can be integrated into the interaction map. Except generating interaction map, the functional classifications of HCIPs also need to be analyzed. Interactors can be grouped by protein domains, molecular functions, biological processes, and signal

pathways, which may help discover common mechanism underlying the proteins of interest. To figure out the new interactions in database, several protein–protein interaction databases such as BioGRID, STRING, IntAct, and MINT can be used to identify the known interaction. However, protein interactions in new publication will not be included in these databases. The interaction information is also not completed, and many known interactions may not be found in these database. Therefore, it is important to dig out protein interaction information in curated literature. Take together, all AP-MS data must be interpreted with care and validated with additional experiments. As with any screening approach, the database does not represent a final or complete interaction network.

2.5 Perspectives

Understanding how proteins interact in complex and dynamic networks is the key to dissect the complexity of many genotype-to-phenotype relationships. The systematic mapping of physical interactions is therefore critical for post-genomic research. Comprehensive analysis of protein–protein interactions is still a challenging endeavor of functional proteomics. Since intrinsic negatives are inherent to every technique, the physical interaction data generated by AP-MS may carry many false positives and negatives. Thus, AP-MS is unlikely to grasp the entire interactome. It is also still a challenge to develop optimal computational tools to visually and computationally represent the multiple layers of data and integrate existing biological knowledge and functional data in literature with the interactome data. Since most AP-MS data represent static graph of PPI map, advanced methods have to be developed and focused on dynamic and spatial changes in PPI.

We have presented the general principles of the AP-MS approach and highlighted some recent developed technologies and successful applications on various signaling pathways. Despite of the increasing AP-MS data and analysis tools, there are still many major challenges. It includes (1) the specificity of protein complex in different cells and tissues, (2) the dynamics of protein complex with different stimulations or posttranslational modifications, (3) the absolute and relative quantitation of proteins, (4) mapping of transient or weak PPI and endogenous PPI from native cells and tissues, (5) the integration of PPI data sets with the other functional data sets, (6) the standardization and benchmarking for interactome mapping, and (7) the challenges for primary cells like neuronal cells and the detection of weak endogenous interaction. Given the different types of mass spectrometric instrumentation, ionization processes, and software platforms, the assessment of published data becomes increasingly difficult. To facilitate sharing experimental data, common standards in data acquisition, data interpretation, and data storage are required.

Many processes in a cell depend on PPI, and perturbations of these interactions can lead to diseases. Comprehensive knowledge of PPI network of signaling pathways will not only give us insights on how the cells respond to stimulation but will also provide new drug targets for therapeutic application. Moreover, many viral

and bacterial pathogens rely on host PPIs to survive in host cells and tissues and exert their damaging effects. Ultimately, such high-quality PPI networks will become invaluable resources for better understanding the mechanisms underlying major human diseases and will enable the better definition of drug targets.

References

Bandyopadhyay S, Chiang CY, Srivastava J, Gersten M, White S, Bell R, Kurschner C, Martin CH, Smoot M, Sahasrabudhe S, et al. A human MAP kinase interactome. Nat Methods. 2010;7(10):801–5.

Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. Science. 2005;307(5715):1621–5.

Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. Nature. 2010;466(7302):68–76.

Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, Cruciat C, Eberhard D, Gagneur J, Ghidelli S, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol. 2004;6(2):97–105.

Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin ZY, Breitkreutz BJ, Stark C, Liu G, et al. A global protein kinase and phosphatase interaction network in yeast. Science. 2010;328(5981):1043–6.

Christianson JC, Olzmann JA, Shaler TA, Sowa ME, Bennett EJ, Richter CM, Tyler RE, Greenblatt EJ, Harper JW, Kopito RR. Defining human ERAD networks through an integrative mapping strategy. Nat Cell Biol. 2012;14(1):93–105.

Colland F, Jacq X, Trouplin V, Mougin C, Groizeleau C, Hamburger A, Meil A, Wojcik J, Legrain P, Gauthier JM. Functional proteomics mapping of a human signaling pathway. Genome Res. 2004;14(7):1324–32.

da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, et al. Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol. 2007;3:89.

Eyckerman S, Verhee A, der Heyden JV, Lemmens I, Ostade XV, Vandekerckhove J, Tavernier J. Design and application of a cytokine-receptor-based interaction trap. Nat Cell Biol. 2001;3(12):1114–19.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989;246(4926):64–71.

Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989;340(6230):245–6.

Fromont-Racine M, Rain JC, Legrain P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat Genet. 1997;16(3):277–82.

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006;440(7084):631–6.

Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, et al. A protein interaction map of *Drosophila melanogaster*. Science. 2003;302(5651):1727–36.

Glatter T, Schittenhelm RB, Rinner O, Roguska K, Wepf A, Junger MA, Kohler K, Jevtov I, Choi H, Schmidt A, et al. Modularity and hormone sensitivity of the *Drosophila melanogaster* insulin receptor/target of rapamycin interaction proteome. Mol Syst Biol. 2011;7:547.

Guerrero C, Milenkovic T, Przulj N, Kaiser P, Huang L. Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. Proc Natl Acad Sci U S A. 2008;105(36):13333–8.

Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O, et al. A protein complex network of *Drosophila melanogaster*. Cell. 2011;147(3):690–703.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999;17(10):994–9.

Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem. 1991;63(24):1193A–203A.

Hirosawa M, Hoshida M, Ishikawa M, Toya T. MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming. Comput Appl Biosci. 1993;9(2):161–7.

Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature. 2002;415(6868):180–3.

Hu P, Janga SC, Babu M, Diaz-Mejia JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, et al. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. PLoS Biol. 2009;7(4):e96.

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A. 2001;98(8):4569–74.

Jager S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K, et al. Global landscape of HIV-human protein complexes. Nature. 2012;481(7381):365–70.

Kerppola TK. Visualization of molecular interactions using bimolecular fluorescence complementation analysis: characteristics of protein fragment complementation. Chem Soc Rev. 2009;38(10):2876–86.

Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, et al. The IntAct molecular interaction database in 2012. Nucleic Acids Res. 2012;40(Database issue):D841–6.

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature. 2006;440(7084):637–43.

Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, et al. Proteome organization in a genome-reduced bacterium. Science. 2009;326(5957):1235–40.

Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, et al. A map of the interactome network of the metazoan *C. elegans*. Science. 2004;303(5657):540–3.

Li ST, Dorf ME. Optimization and ZSPORE analysis of affinity aurification aoupled with tandem mass spectrometry in mammalian cells. J Proteomics Genomics Res. 2013 (in press).

Li S, Wang L, Berman M, Kong YY, Dorf ME. Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. Immunity. 2011;35(3):426–40.

Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E, et al. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. 2012;40(Database issue):D857–61.

MacBeath G. Protein microarrays and proteomics. Nat Genet. 2002;32(Suppl):526–32.

MacCoss MJ, Wu CC, Yates 3rd JR. Probability-based validation of protein identifications using a modified SEQUEST algorithm. Anal Chem. 2002;74(21):5593–9.

Mirgorodskaya OA, Korner R, Kozmin YP, Roepstorff P. Absolute quantitation of proteins by acid hydrolysis combined with amino acid detection by mass spectrometry. Methods Mol Biol. 2012;828:115–20.

Mueller CL, Jaehning JA. Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. Mol Cell Biol. 2002;22(7):1971–80.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1(5):376–86.

Parrish JR, Gulyas KD, Finley Jr RL. Yeast two-hybrid contributions to interactome mapping. Curr Opin Biotechnol. 2006;17(4):387–93.

Pfefferle S, Schopf J, Kogl M, Friedel CC, Muller MA, Carbajo-Lozoya J, Stellberger T, von Dall'Armi E, Herzog P, Kallies S, et al. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. PLoS Pathog. 2011;7(10):e1002331.

Pilot-Storck F, Chopin E, Rual JF, Baudot A, Dobrokhotov P, Robinson-Rechavi M, Brun C, Cusick ME, Hill DE, Schaeffer L, et al. Interactome mapping of the phosphatidylinositol 3-kinase-mammalian target of rapamycin pathway identifies deformed epidermal autoregulatory factor-1 as a new glycogen synthase kinase-3 interactor. Mol Cell Proteomics. 2010;9(7):1578–93.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods. 2001;24(3):218–29.

Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, Armstrong CM, Li S, Jacotot L, Bertin N, Janky R, et al. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nat Genet. 2003;34(1):35–41.

Rossi F, Charlton CA, Blau HM. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. Proc Natl Acad Sci U S A. 1997;94(16):8405–10.

Rozen R, Sathish N, Li Y, Yuan Y. Virion-wide protein interactions of Kaposi's sarcoma-associated herpesvirus. J Virol. 2008;82(10):4742–50.

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005;437(7062):1173–8.

Smith BE, Hill JA, Gjukich MA, Andrews PC. Tranche distributed repository and ProteomeCommons. org. Methods Mol Biol. 2011;696:123–45.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011;27(3):431–2.

Sowa ME, Bennett EJ, Gygi SP, Harper JW. Defining the human deubiquitinating enzyme interaction landscape. Cell. 2009;138(2):389–403.

Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, et al. The BioGRID Interaction Database: 2011 update. Nucleic Acids Res. 2011;39(Database issue):D698–704.

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011;39(Database issue):D561–8.

Tagwerker C, Flick K, Cui M, Guerrero C, Dou Y, Auer B, Baldi P, Huang L, Kaiser P. A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivo cross-linking. Mol Cell Proteomics. 2006;5(4):737–48.

Tewari M, Hu, PJ, Ahn JS, Ayivi-Guedehoussou N, Vidalain PO, Li S, Milstein S, Armstrong CM, Boxem M, Butler MD, et al. Systematic interactome mapping and genetic perturbation analysis of a C. elegans TGF-β signaling network. Mol Cell. 2004;13:469–82.

Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res. 2003;31(1):334–41.

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature. 2000;403(6770):623–7.

Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, Rajagopala SV, Roupelieva M, Rose D, Fossum E, et al. Herpesviral protein networks and their interaction with the human proteome. Science. 2006;311(5758):239–42.

Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science. 2000;287(5450):116–22.

Waugh DS. Making the most of affinity tags. Trends Biotechnol. 2005;23(6):316–20.

Shitao Li, Ph.D., USA Shitao Li is a research fellow in Department of Microbiology and Immunobiology, Harvard Medical School. He obtained his Ph.D. from Wuhan University, China. He is a recipient of Kaneb Fellowship and AAI Abstract Award (2010). Dr. Li studies on protein interaction network using proteomics approach. He has mapped a dynamic antiviral innate immunity protein interaction network and currently is working on virus-host protein interaction network. By examining the protein network, he is investigating the signaling mechanisms controlling innate antiviral immunity and new drug targets for host defense to viral infection. He published his research on several prestigious journals such as Nature, Immunity, and Molecular Cell.