UC Irvine
UC Irvine Previously Published Works

Title

From pathways to networks: Connecting dots by establishing protein–protein interaction networks in signaling pathways using affinity purification and mass spectrometry

Permalink
https://escholarship.org/uc/item/05m852fm

Journal
Proteomics, 15(2-3)

ISSN
1615-9853

Authors
Li, Xu
Wang, Wenqi
Chen, Junjie

Publication Date
2015

DOI
10.1002/pmic.201400147

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

HHS Public Access

Author manuscript

Proteomics. Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:

Proteomics. 2015 January ; 15(0): 188–202. doi:10.1002/pmic.201400147.

From pathways to networks: connecting dots by establishing protein-protein interaction networks in signaling pathways using affinity purification and mass spectrometry

Xu Li, **Wenqi Wang**, and **Junjie Chen**[*]

Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX77030, USA

Abstract

Signal transductions are the basis of biological activities in all living organisms. Studying the signaling pathways, especially under physiological conditions, has become one of the most important facets of modern biological research. During the last decade, mass spectrometry has been used extensively in biological research and is proven to be effective in addressing important biological questions. Here, we review the current progress in the understanding of signaling networks using mass spectrometry approaches. We will focus on studies of protein-protein interactions that use affinity purification followed by mass spectrometry approach. We discuss obstacles to affinity purification, data processing, functional validation, and identification of transient interactions and provide potential solutions for pathway-specific proteomics analysis, which we hope one day will lead to a comprehensive understanding of signaling networks in humans.

Keywords

affinity purification; Hippo signaling; mass spectrometry; protein-protein interaction; signaling pathways

Application of mass spectrometry in the studies of signaling pathways

Sophisticated signaling networks are required for development and survival, and a minor disruption of any of these pathways may cause severe diseases such as cancer in humans. A typical signaling pathway, such as transforming growth factor–beta (TGF-β) signaling, usually starts with an extracellular ligand (e.g., TGF-β) binding to a receptor (e.g., TβR) on the cell membrane. This ligand-receptor interaction could trigger a sophisticated phosphorylation cascade, followed by translocation of major signaling molecules (e.g., Smads) to the nucleus and binding to specific DNA sequences to regulate gene expression and thus ultimately modulate cellular activities (reviewed in [1]). Over the years, considerable efforts have been devoted to elucidating signaling transduction pathways in human cells. A number of pathways have been identified, in which protein-protein

[*]Corresponding author: Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard - Unit 66, Houston, Texas 77030, Ph: 713-792-4863, Fx: 713-745-6141, jchen8@mdanderson.org.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

interactions (PPIs) and post-translational modifications (PTMs) are often involved. As a quantitative and unbiased approach, mass spectrometry (MS) has been proven to be a powerful tool in revealing protein abundance as well as PPIs and PTMs under near-physiological conditions (Fig. 1) (reviewed in [2, 3]). There are many decision points when using MS to study signaling events, which include sample preparation, choice of MS applications, pre-MS strategies, the MS itself, and post-MS data analysis (Fig. 1).

Signaling events are by definition associated with changes, for example the change of protein level inside or outside of the signaling cascade, which determine the final outcome. Although many approaches, such as microarray, may be used to detect mRNA level changes, protein levels are frequently regulated at post-transcriptional and post-translational levels and therefore do not always correlate with mRNA levels. Monitoring protein abundance, especially in a complex environment, provides irreplaceable insights into signaling events. Using a “shotgun” strategy, proteins can be directly identified from a complex mixture by enzymatic digestion followed by reverse-phase chromatography and tandem mass spectrometry (MS/MS) [4]. Because of its simplicity, monitoring signal-induced protein level changes, such as those that occur following the binding of an upstream ligand or treatment with an inhibitor, has become one of the most widely employed signaling-related applications of MS. To allow measurement of the absolute or relative protein abundance among different samples, tremendous efforts have been made to ensure that MS results are more reliable and accurate from a quantitative point of view. These efforts began with the traditional semi-quantitative MALDI-TOF and liquid chromatography (LC)-MS/MS (MS2) approach (reviewed in [2, 5]), to recently developed isotope-coded affinity tags (ICAT) [6], stable isotope labeling by amino acids in cell culture (SILAC) [7], isobaric tags for relative and absolute quantification (iTRAQ) [8], tandem mass tags (TMTs) [9, 10] and triple-stage mass spectrometry (MS3) [11]. With these technological advancements, measurements of protein relative and absolute abundance become more robust, precise and accurate. For example, the TMT technology allows as many as 54 samples to be tagged with different combination of isobaric tags and analyzed in a single MS run, thereby providing relative protein abundance in each sample [12–14]. iTRAQ and TMT do not require growing cells in isotope-containing culture medium, therefore they could be used for comparing protein abundance in multiple tissue samples [15]. However, it has been frequently observed that the accuracy and precision of quantitative data from iTRAQ and TMT suffer because the contaminating near-isobaric ions are isolated and fragmented together with the target ions [16, 17]. Efforts have been made to solve this problem, which include the use of MS3 and improving data analysis [11, 16]. Indeed, the technological developments often require accompanying advances in bioinformatics methodologies. A handful of software packages, such as OpenMS [18, 19] and MaxQuant [20, 21], were designed to analyze the label-free or isotope-labeled MS data. Specific software packages, such as virtual expert mass spectrometrist (VEMS), were also designed specifically for iTRAQ or TMT data analysis, to achieve more accurate peptide ratios [22, 23].

These new MS-based technologies provide great advantages in detecting differences in protein abundance in signaling events between different states, such as basal versus activated or inhibited, or healthy versus diseased. Using these technologies, researchers have

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

been able to compare protein abundance in complex environments, such as haploid versus diploid yeast [24] or healthy tissues versus tumor samples [25]. With these quantitative methods in hand, ligand-induced receptor signaling pathways, which can be easily manipulated by external ligand stimulation or treatment with inhibitors, are now being studied. For example, using a γ-secretase inhibitor to block Notch signaling, Dai *et al*. profiled dose-dependent protein expression in glioblastoma cancer stem cells [26]. Treating Madin-Darby Canine Kidney Epithelial cells with TGF-β ligand, Shen *et al*. profiled the plasma membrane proteins and found Wnt-5a involved in H-Ras/TGF-mediated epithelial-mesenchymal transition [27]. With a similar strategy, Kim *et al*. profiled the total protein changes on lipid rafts upon TGF-β1 treatment [28]. Using an iTRAQ-based quantitative screen of Smad4-negative and stably rescued SW480 cells, Ali *et al*. found a series of protein level differences between the two cell lines, which indicates that they could be potential downstream targets of the TGF-β signaling pathway [29].

MS-based approaches also provide insight into PTMs that regulates protein activity and stability. Proteome-scale studies have been performed to assess the global change of PTMs, such as phosphorylation (reviewed in [30–32]), ubiquitination and Ub-like modifications (reviewed in [33, 34]), acetylation (reviewed in [35]), methylation (reviewed in [36]), and glycosylation (reviewed in [37]). Among many PTMs, phosphorylation is one of the most important PTMs in signaling transduction, and phosphoproteomics is the most studied PTM using MS approaches. Functionally focused large-scale proteomics studies have also been conducted, such as studies of DNA damage–induced phosphorylation by Ataxia telangiectasia mutated (ATM) and Ataxia telangiectasia and Rad3 related (ATR) [38] and of cell cycle regulations [39]. With proper enrichment strategies and quantitative MS approaches, several signaling pathways have now been characterized using MS-based global phosphoproteomic profiling (reviewed in [32]), such as quantitative phosphoproteomics of TGF-β signaling [40], Wnt signaling [41], insulin signaling [42] and proto-oncogene tyrosine-protein kinase Src signaling [43]. Affinity purification (AP)/MS also has been used frequently in enriching and identifying PTMs of important proteins in signaling pathways. Willert *et al*. purified active Wnt molecules and found that they are lipid modified and act as stem cell growth factors [44]. Morén *et al.* identified K507 on Smad4 as a ubiquitination site and participates in R-Smad C-terminal phosphoserine recognition [45]. Matsuura *et al.* found O-GlcNAc modification occurs at multiple sites on Notch epidermal growth factor repeats [46]. Together with the extensive experimental studies, the bioinformatics tools for analyzing protein PTMs have also been rapidly developed. The MS results can be motif analyzed [47], Gene Ontology (GO) and pathway annotated [48, 49], and PPI database incorporated [50] and eventually provide a comprehensive understanding of cellular events under physiological and pathological conditions (reviewed in [51]). These studies provide insights into the regulation of signaling pathways and are valuable resources for basic and clinical research.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Affinity purification followed by mass spectrometry – a powerful tool in identifying protein-protein interactions in signaling pathways

Studies of PPIs have provided immense insights into protein functions. Many biochemical tools have been invented to explore PPIs, including *in vivo*/*in vitro* pull-down assays, co-immunoprecipitation (co-IP) assays using epitope tags or antibodies against endogenous proteins, crosslinking protein interaction assays and/or label transfer protein interaction analysis (e.g., biotinylation labeling), and yeast two-hybrid analysis (reviewed in [52]). As a relatively new and unbiased approach, AP/MS offers tremendous advantages over other methods in identifying PPIs under near-physiological conditions and identifying protein complexes instead of binary interactions [2]. By performing AP of a protein of interest (the "bait"), and following it with LC-MS/MS, the partner proteins (the "preys") that form complexes with the bait can be identified [53]. We have used this approach to study the DNA damage signaling pathways (for examples, see [54–58]; for a review, [59]), since many stable and functionally relevant complexes are formed before and after DNA damage or regulated in a cell cycle–dependent manner. It helped us tremendously, since many insights into the regulation of DNA damage–responsive pathways were initiated from the AP/MS results of individual DNA damage–responsive proteins.

Using AP/MS to study individual proteins in other signaling pathways, such as the TGF-β signaling [60, 61] and Wnt signaling [62] pathways, has also led to many important findings in these fields. With three different MS strategies, Luo *et al*. profiled Smad2-interacting proteins under different TGF-β stimulation conditions [61]. Several other core components in TGF-β signaling, such as Smad3 [63], Smad4 [60], TβRI [64], and TGFβ1 [65], have also been studied by AP/MS. Using a streptavidin-based tandem AP (TAP)/MS approach, Angers *et al*. and Major *et al*. obtained the interactomes of seven core components of the Wnt signaling pathway, including Dishevelled-1/2/3 (DVL1/2/3) [66], β-catenin, AXIN1, adenomatous polyposis coli (APC), and beta-transducin repeat containing E3 ubiquitin protein ligase 1/2 (β-TRCP1/2) [62]. These studies uncovered several novel and important components of Wnt regulators, such as Cullin-3 complex and Wilms tumor gene on X chromosome (WTX) [62, 66]. Using a triple SILAC technology, Hilger *et al*. also investigated APC, Axin-1, DVL2, and C-terminal binding protein 2 (CtBP2) interactomes under various stimuli and detected several stimulus-dependent interactions such as the binding of APC with Girdin and Axin-1 [67].

Most studies, however, focused only on one or few components in a given pathway. For example, SMAD4 and β-catenin, the respective core components of the TGF-β and Wnt signaling pathways, have been affinity purified by many groups using a variety of methods, while many other important components, such as other Smads in TGF-β pathway, the T-cell factors/lymphoid enhancer-binding factor 1 (TCFs/LEF1), LDL receptor–related protein (LRPs) and Frizzled (FZDs) in Wnt pathways have never or rarely been studied. Moreover, the data generated by different groups using diverse AP/MS strategies and filtration criteria are not comparable. Therefore, it remains challenging to integrate different protein AP/MS results into a unified network for the entire signaling pathway. Systematic investigation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

using a unified approach is urgently needed to provide a network view of PPIs involved in these signaling pathways.

Using AP/MS in large-scale proteomics studies – methodology, data analysis, and validation

Several large-scale AP/MS-based studies, consisting of hundreds to thousands of protein purifications, have been conducted in yeast [68–70], drosophila [71] and humans [72] and uncovered thousands of PPIs. These studies, as well as yeast two-hybrid screenings performed with baits from many model organisms [73, 74], have been used to build huge interactomes containing millions of PPIs. Since protein-protein interactions imply functional connections between these proteins, these PPI data have become invaluable resources for further investigating functions of individual proteins. Moreover, these large-scale studies provide a broad view of the whole genome complexomes. However, to get a better understanding of certain biological process, more sophisticated and focused study is a must, since it allows in-depth functional validation. Several function-related medium-to-large-scale AP/MS studies in human cells have been conducted in recent years, which focus on a group of proteins with similar functions, such as chromatin remodeling complexes, deubiquitinating enzymes (DUBs), autophagy systems, and histone deacetylase (HDAC) families [75–78]. Some of these studies were coupled with siRNA screenings and functional validations [78]. These focused proteomics studies have uncovered many novel PPIs and protein complexes and provided numerous leads for further functional studies of these cellular processes.

Methodology

There are four major steps in AP/MS: bait presentation, AP, MS, and data analysis [2]. First, to perform the AP, there must be a way to bring down the bait from the cell lysates. Ideally, antibodies targeting endogenous proteins are the best choice because they cause minimal disturbance to the system. Besides their high cost, however, another major drawback of using antibodies against endogenous proteins in large-scale MS is that they often bring down nonspecific cross-reactive binding proteins, which are impossible to eliminate through bioinformatics analysis and are detrimental to the quality of large-scale datasets. Despite these technical difficulties, an exciting attempt using over 3,000 antibodies against endogenous proteins in human cells has led to the discovery of many novel protein complexes [72].

As an alternative approach, epitope tags are frequently used for AP in large-scale proteomics studies [2, 52, 53]. In yeast, efficient homology recombination made it possible to introduce tandem epitope tags to the endogenous gene locus [68, 70]. In mammalian cells, however, this approach is technically challenging and not suitable for large-scale studies [5]. Instead, the tagged proteins are usually introduced into cells through transfection or viral infection. Transiently transfected cells can be directly subjected to AP/MS, but that usually leads to the bait being expressed at relative high levels. Since overexpressed proteins often bind to proteins that they normally do not interact with under physiological conditions, the transient transfection followed by AP/MS approach may lead to considerable false-positive findings.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A better approach may be to establish stable pools or cell lines with an expression level of bait protein similar to that of the endogenous protein [79].

One-step AP/MS using different epitope tags such as FLAG peptide tag (FLAG), human influenza hemagglutinin peptide tag (HA), and Green fluorescent protein (GFP) is the most frequently used approach in medium-to-large-scale PPI studies in mammalian cells [76, 78, 80]. The biggest advantages are obvious: it is economically and technically convenient and is able to generate a long list of interacting proteins [79]. Another major advantage is, since it has been used extensively in large-scale studies of mammalian cells and produced many publicly available datasets, the background generated by certain tags, such as FLAG-tag, is well known and is relatively easy to eliminate. With the quick development of MS equipment, advanced mathematical modeling, and data-driven elimination of common contaminants, one-step AP now provides more reliable results than it once did. In the recent Hippo pathway studies, AP/MS analysis using FLAG or 3XFLAG tag were able to generate dependable results with very few previously known contaminants such as keratin and chaperone proteins [80, 81].

Tandem affinity purification usually consists of two or more AP steps against different epitope tags fused to the bait protein. The original version of TAP tags contained a tobacco etch virus (TEV) protease cleavage site between two different tags, which could be cleaved by TEV protease and the protein complex released from the beads and subjected to the second step of AP [82]. Designed for protein complex characterization and proteome exploration [82] and first used in large-scale yeast proteomics studies [68, 70], TAP now has over twenty different modified versions with different combination of tags and has become the standard approach in several model organisms [83, 84], but not yet in humans. The major reasons are that it is relatively costly and technically challenging at the cleavage step during AP, hindering its adaptation for large-scale proteomics studies in human systems.

However, high-throughput MS/MS is also time consuming and very expensive. Significant efforts have been wasted on repeatedly re-identifying the same abundant peptides or peptides from “sticky proteins” that bind to many baits in AP/MS analysis [85]. The presence of these abundant and common contaminants may lead to the low coverage of peptides from specific prey proteins. Thus, the real signals are often buried in the noisy background, such that several technical repeats and sophisticated equipment are needed to achieve a high percentage of coverage for specific preys [86]. Since TAP or its modified versions eliminate a large fraction of abundant proteins during the two-step AP, it is potentially a more effective way of detecting “regulated” PPIs. Because these regulated interactions are critical for studies of signal transduction pathways, the TAP/MS approach may provide some level of advantage in medium-to-large-scale studies focusing on a particular signaling pathway. However, the most technically challenging part of TAP/MS is the cleavage/elution step. In order to establish PPI networks that are useful for directing in-depth functional studies in mammalian systems, a more user-friendly TAP method similar to those used in yeast proteomic studies is critically needed. Moreover, a data-driven computational method specifically designed for the TAP-MS approach would also help to uncover many *bona fide* interacting proteins, but at the same time eliminate contaminants and abundant proteins that are often associated with the baits in human systems.

Author Manuscript

In addition to traditional AP/MS or TAP/MS, quantitative AP/MS is another way to capture signal-induced changes in interactomes. Similar to its advantage in monitoring signal-induced whole proteome changes and PTM changes, labeled quantitative AP/MS approaches such as SILAC or iTRAQ also offer the ability to quantitatively compare between different stages in signaling events. SILAC based AP/MS has been used in several small-medium scale studies [67, 87]. For example, using SILAC to differentially label proteins in epidermal growth factor (EGF)-stimulated versus unstimulated cells identified proteins regulated by EGF signaling and their associated partners [88, 89]. The quantitative AP/MS approach holds lots of promise for future studies, but it is still technically challenging and available only in a limited capacity.

Data analysis, visualization and validation

Considerable efforts were devoted to MS data analysis (reviewed in [3, 90]). Raw spectra need to be processed first. The processing steps include noise filtering, background subtraction, isotope and charge deconvolution, peak detection, mass calibration and retention time alignment. Processed MS/MS spectra are searched against protein sequence databases thereby assigning MS/MS spectra to peptides, which can subsequently be mapped to proteins (reviewed in [3]). Target-decoy search strategy is usually used for increasing confidence and estimating false discovery rate (FDR) [91, 92]. The frequently used data processing software include SEQUEST [93], MASCOT [94] and X!Tandem [95].

Then the assigned PPIs are needed to be further analyzed to determine the *bona fide* interactions from background noises. Currently there are at least three well-accepted data filtration methods for label-free shotgun MS data analysis, all of which are based on the spectra counts from AP/MS results. These are the normalized spectral abundance factor (NSAF) method [75] and its data-driven expansion based on the BCM-CCI algorithm [72]; the Comparative Proteomic Analysis Software Suite (CompPASS) [76]; and the Significance Analysis of INTeractome (SAINT) method [96]. The NSAF algorithm was one of the first computational methods for this application, which normalizes the raw spectra counts with prey protein length and total spectra counts. This normalization idea has been adopted for other methods. The probability of interactions within and between complexes is computed solely on the basis of NSAFs using the Bayes approach [75]. Malovannya *et al.* established a data-driven version based on this method, which utilizes their large-scale proteomics data and is able to assign scores to each of the interactions [72, 97].

CompPASS was originally designed for analyzing the large-scale DUB PPI dataset [76]. Starting with two scores, Z and D, which are, respectively, the same as the Z-score used in statistics and the score representing prey specificity/data reproducibility, they introduced another score, WD, in their follow-up studies [78], which weighted the D-score to distinguish the “background proteins that were known to be true interactors” for specific baits from “background proteins that were never true interactors.” It largely relies on data reproducibility and has been proven to be effective with different datasets [98].

The SAINT method was originally designed for analyzing the yeast kinase and phosphatase interactomes obtained from yeast two-hybrid data [99]. This method has now been proven to be effective in the analysis of AP/MS data [80, 100, 101]. The SAINT method also uses

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

advanced mathematical models to estimate the abundance and specificity parameters and therefore is capable of generating probability scores for each of the interactions. Several follow-up studies have been published, including SAINT-MS1, dealing with MS1 intensity data [102]; SAINT express, with a simpler statistical model [103]; and contaminant repository for affinity purification–mass spectrometry data (CRAPome), dealing with the common contaminants [104].

Several other modified data-driven algorithms have been used in individual studies [71, 105]. These computational algorithms could also be used together, or even combined with hand curating, to generate the list of high-confidence candidate interacting proteins (HCIPs) [81, 105].

Once the HCIPs have been identified and the interaction networks built up, Cytoscape software is usually used to visualize the interacting networks [106, 107]. With many plug-ins (i.e., apps), Cytoscape provides even more powerful features, such as topological analysis, clustering, functional enrichments, network comparison, and merging [108]. Other “R” based visualization methods are also available. The R framework and Bioconductor [109], are also frequently used for data analysis, clustering, and visualization. GO and pathway annotation [48, 49] are usually carried out using the HCIPs to achieve an understanding of pathway interactions.

For data validation, co-IP assay using epitope tags or antibodies against endogenous proteins is the most frequently used approach [78, 80, 81, 100, 101]. Other approaches include two-hybrid– and flowcytometry– based interaction validations, such as Fluorescence resonance energy transfer (FRET) and Bimolecular fluorescence complementation (BiFC) assays [110–113]. Reciprocal AP/MS is occasionally used for a few preys of interest, with a goal of recovering the original bait from AP/MS using its prey as the bait [77, 100]. Another part of the validation is to use the interaction data and search them in various PPI databases, such as STRING [114], BioGrid [115], Intact [116], BIND [117], DIP [118], and HPRD [119], to show that the newly identified interactomes at least partially overlap with these knowledge databases. The HCIPs could also be searched in the NetPath database [120], a resource of functional protein complexes in many signal transduction pathways, to get an idea of coverage of known proteins in the given pathway. In-depth functional validation, sometimes coupled with RNAi screening, could be conducted for several proteins of interest to prove that the dataset can lead to biological meaningful results [100, 101]. A few examples of data validation could be found in the following section of Hippo pathway studies. The whole strategy of AP/MS has been summarized and illustrated in Fig. 2.

Systematic proteomics studies of Hippo signaling pathway – an example of using AP/MS in the study of signal transduction pathways

There are many advantages to systematic analysis of a signaling pathway using medium-to-large-scale AP/MS, which purifies the core components or all of the known proteins involved in a given pathway. New components or regulators could be identified through AP/MS, which can be further functionally validated [2]. Comparing the purification results under different conditions or following treatment with inhibitors or drugs may also gain

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

additional insights into protein functions and regulations. Several medium-to-large-scale proteomics studies focused on certain signaling pathways have been conducted; for example, soon after the large-scale TAP/MS studies in yeast [68], Bouwmeester *et al*. performed TAP/MS for 32 known and candidate proteins in the human TNF-α/NF-κB pathways and identified 80 novel interactors [121]. Glatter *et al*. studied the InR/TOR pathway in *Drosophila* using AP/MS and identified 58 new network components from 16 known baits in the pathway [122]. These studies provided new insights into these pathways and identified potential novel components of these pathways, which significantly benefited the research in these fields.

Recent studies by four different groups using AP/MS methods for investigating the Hippo-YAP signaling pathway, an important pathway that controls organ size and tumorigenesis, provide an interesting example of the variation in methods for using AP/MS approach in studying signal transduction pathways [80, 81, 100, 101]. Early components of the Hippo pathway, including *warts* (*wts*) [123, 124], *salvador* (*sav*) [125, 126], *hippo* (*hpo*) [127, 128], mob as tumor suppressor (*mats*) [129], *yorkie* (*yki*) [130], *merlin* (*mer*), and *expanded* (*ex*) [131], all of which are involved in organ size control, was identified from genetic mosaic screens in *Drosophila* (reviewed in [132]). Later, the mammalian homologs of these proteins, including Lats1/2, WW45, Mst1/2, Mob1, YAP/TAZ, NF2, and FRMD6, were identified and shown to function in a similar signaling pathway (Fig. 3A) (reviewed in [133]). Although these known components have been extensively studied during the last decade, the identification of new components in the Hippo pathway has been relatively slow. Over 30 components of the human Hippo pathway have been identified during the last decade. However, compared with other well-defined signaling pathways such as Wnt and MAPK that are also known to be involved in tumorigenesis, very few mutations have been discovered in known Hippo pathway genes in common human cancers. Neurofibromin 2 (*NF2*) is the only Hippo pathway gene known to be frequently mutated and inactivated in cancers [134]. This suggests that there might be undefined Hippo pathway components that are mutated and responsible for tumorigenesis in some cases. Alternatively, the regulation of the Hippo pathway, presumably through PTMs and PPIs, may be more relevant to human carcinomas than the somatic mutations of genes involved in this pathway. In either case, identifying new components and new regulatory mechanisms of this pathway is essential for further study of its role in organ size control and human cancers.

Using a modified GS-TAP system designed for protein purification in *Drosophila* [135, 136], Kwon *et al*. performed TAP/MS for 12 core components of the Hippo pathway in *Drosophila* and identified a high-confidence PPI network consisting of 153 HCIPs [100]. They conducted co-IP validation for 26 interactions using epitope-tagged proteins, and 23 of them were positive. They further carried out an RNAi screening using 255 dsRNAs targeting the 153 HCIPs and monitored the effect on this signaling pathway using a Yki-reporter assay. Seventy-seven negative and 23 positive regulators were identified that showed >1.5-fold change of Yki-reporter activity. They subsequently performed reciprocal TAP/MS using Yki-interacting protein CG4674 (leash) and recovered Yki, which indicates that at least some of these interactions occur under physiological conditions. Finally, they functionally validated this protein leash as a Yki regulator in *Drosophila*, since

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

overexpression of leash led to significant reduction in wing size [100]. This work provides a standard workflow for future proteomics studies of signaling pathways, which start with AP/MS, followed by interactome validation and further confirmed by functional validation.

Similarly, Wang *et al*. performed streptavidin-FLAG-S (SFB) triple-tagged TAP/MS for 32 Hippo pathway components in human cells. They identified 343 HCIPs and functionally validated WWC1-interacting protein CIT and YAP1-interacting protein CCDC85C as new Hippo pathway components that regulate YAP1 subcellular localization and activities [101]. Couzens *et al*. performed 3xFLAG-tagged AP/MS coupled with proximity-dependent biotin identification (BioID) AP/MS [137] of 27 Hippo pathway components treated or not treated with serine/threonine phosphatase inhibitor okadaic acid and identified 400 HCIPs [80]. Since phosphorylation cascades are often involved in signal transductions, the interactomes of the Hippo PPI networks of treated and untreated cells could be compared to suggest putative mechanisms for the regulation of this pathway and further functional implications [80]. In addition, Hauri *et al*. performed AP/MS for 34 bait proteins related to the human Hippo pathway (8 core components, 26 secondary baits) and identified 270 HCIPs [81]. Taken together, the findings from these four groups uncovered nearly one thousand HCIPs, which should be considered new potential components of the Hippo pathway (Fig. 3B). It is likely that these newly identified components will be the starting points for further study of the regulation of the Hippo-YAP signaling pathway and how it may be connected with other cellular processes and signaling pathways. Moreover, some of these new components may serve as potential biomarkers or drug targets for cancer diagnosis and treatment in the future.

Current obstacles for the AP/MS-based large-scale studies of signal transduction pathways

Unfortunately, there are still some conceptual and technical challenges to proteomics studies of signaling pathways. While current AP/MS methods are effective in identifying stable protein complexes, they are inadequate in recognizing regulated interactions, which are essential for understanding the complex signaling networks in the cell. This shortcoming is particularly noticeable in large-scale proteomics studies, since the appearance of abundant associated proteins in the interactomes drastically reduces the sensitivity for detecting low levels of regulated but biologically significant interactions. Doing AP/MS in different organism, conceptual biases, and differences in AP approaches all lead to very distinct results. Since different organisms are genomically different, the results from different organisms are usually not comparable [138]. However, proteins encoded by genes mutated in inherited genetic disorders are likely to interact with proteins known to cause similar disorders, which suggest the existence of disease PPI subnetworks [138]. This “guilty by association” concept may also apply to signaling pathways, since many signal transduction pathways, such as Wnt and Hippo, are highly conserved across species and their corresponding PPI networks should also display a high degree of similarity.

Another technical challenge is the diversity of AP methods that are being used in the field, which makes it difficult to compare results from different investigators. As already discussed, using epitope tags is still the most cost-effective approach for large-scale proteomics studies. This is unlikely to change before unified and quality-controlled

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

collection of antibodies against endogenous proteins becomes available [2, 52, 53]. It was noticed from the very beginning that one-step purification [69] and tandem purifications [68] in yeast generate very different results even with the same baits [139], which highlights the importance of having unified or comparable technical platforms. In yeast, TEV-based TAP has become a standard approach, and many large-scale studies have been performed on this platform [68, 70]. In human studies, however, there are at least thirty different epitope tags or combinations that have been used for AP. In the recent Hippo signaling studies, for example, five different AP methods were used by four groups, of which three were one-step APs and two were modified TAPs. The datasets from the first three groups mentioned above, one in *Drosophila* and two in human, are hardly comparable: less than 10% of the HCIPs were present in all three datasets and only ~20% in the two human datasets [140]. Since these three groups used very similar bioinformatics approaches for data filtration [96], the discrepancies likely come from the AP approaches they used.

It is difficult to conclude which AP method is the best for large-scale proteomics studies until a systematic comparison among different AP methods and painstaking data validation have been conducted. Nevertheless, if it is possible, a unified AP/MS method(s) in human proteomics studies, just like the TEV-based TAP/MS method used in yeast, would provide great advantages in understanding signaling events in humans, since it would permit meaningful comparisons of results generated by many investigators. Interlaboratory collaborations have been made to achieve high reproducibility of large-scale human protein-complex analysis by using a standardized TAP/MS approach with same reagent and data analysis [105]. The hope is that one day the MS data generated by different research groups could become comparable and integrated into one large database, as happened for microarray data. This will greatly advance the field and profoundly change how AP/MS and other proteomics data can be used for functional studies.

On the other hand, one could also argue that technological diversity may in some cases be beneficial and required, since using one unified method could in theory introduce strong technical biases. Tandem affinity purification usually provides “cleaner” results with fewer common contaminants, but it is more technically demanding than one-step AP and may cause loss of transient or weak interactions during the cleavage/elution step [2]. One-step AP, such as the use of FLAG-tag, usually produces a list of potential interacting proteins that is longer than the one obtained from TAP, and it has been used in many milestone studies on a large scale. However, one-step AP has its own technical limitations, such as in detecting transient or regulated interacting proteins in the presence of highly abundant, non-specific associated proteins [141, 142]. Several other AP techniques have been developed recently to capture the weak/transient interactions, such as BioID [137]. The BioID method is able to produce meaningful data that reflect “local protein partners” information, which could not be generated by the traditional AP method [80]. However, it remains to be determined whether this method is able to capture weak or transient interactions, such as enzyme-substrate interactions. Ideally, at least two unified AP methods should become standards in the field for signaling studies: a more stringent method to characterize stable or regulated complexes, and a more sensitive method that captures transient or weak interactions that are functionally important for studies of signal transduction.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Another challenge in the field is data filtration. With the advances in MS technology and bioinformatics, we are gradually overcoming the obstacles of generating raw spectra counts and mapping them to correct proteins [143]. We are now increasingly facing the complication of our own success: how to make sense of a list of hundreds, if not thousands, of putative associated proteins? Before the systematic bioinformatics data filtration approaches were developed, the data were usually manually curated and filtered by experience [2]. That may be feasible for studies of individual proteins, since the number of proteins in one MS list, no matter how long it may be, is still limited, and follow-up confirmation and functional studies are usually carried out to make sure that the right PPIs are presented. For large datasets such as whole pathway proteomes, however, automated data filtration algorithms are needed.

The current data filtration methods work efficiently for their originally designated applications [72, 75, 76, 96]. Methods such as SAINT, which use semi-supervised mixture models and Makov chain Monte Carlo (MCMC) sampling, is mathematically advanced and works effectively with a group of unrelated proteins [96]. Such an assumption may not be true in biology, however, since in many cases a connection between unrelated proteins may lead to new scientific discoveries. In signaling pathway studies, data filtration becomes even more challenging, since proteins often interact with other proteins in the same pathway. Together with their interacting proteins, they frequently appear in other purification results across experiments in the same pathway. This severely interferes with the estimated parameters for abundance and specificity and could lead to both false-positive and false-negative identifications. Such a problem may be difficult to solve solely by mathematical modeling, but customized data-driven mathematical algorithms may offer a solution. Combining high-throughput datasets with mathematical modeling will help greatly in identification of interactions with more confidence and fewer false positives [79]. Several attempts have been already made, such as establishment of the Drosophila Protein Interaction Mapping project (DPIM) [71] and contaminant repository for AP/MS data (CRAPome) [104]. More investigations in AP/MS data processing will undoubtedly advance this quickly growing field.

Future Perspectives

The technological advancement in the field of AP/MS analysis for proteomics studies is far exceeding our ability to handle the information properly. Sometimes even simple questions face a dilemma. For example, AP/MS is often performed with biological replicates and technical replicates; should the not-well-reproducible data (e.g., identified twice in four replicates) or the interactions with low peptide counts/interaction scores be automatically excluded? For studies of individual proteins, the answer should surely be no. Since one AP/MS covers only a limited portion of the total peptide population [86], and some proteins are relatively difficult to recognize by MS because of their abundance or sequence/structural features [53], the peptide readings per se may not reflect the real biological importance of these protein-protein interactions, especially if these interactions are regulated in a signaling pathway. In this case, the functional validation and relevance is the most important aspect and these leads should be pursued with a biological question in mind.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For large-scale proteomics studies, the answer to the same question may be ambiguous. Large-scale datasets focusing on signaling pathways should not only provide a valid resource for the research community but also should generate new insights from an overall view of the biological system, such as functional correlation with different biological processes or other signaling pathways, or even provide new disease correlations based on the interactions identified. Without proper data filtration, the GO process and functional implication may not be meaningful, largely because of the presence of abundant contaminants in the datasets. In these cases, the criteria for data filtration should be stringent to reveal the true biological significance.

On the other hand, large-scale MS studies are very expensive and time consuming. We should utilize any information gained from these studies as fully as possible. Luckily, other large-scale approaches, such as RNAi screening, microarrays, metabolomics, next-generation sequencing, and even patient mutation/disease correlation databases, are available and continually growing. The integration of data generated via different technologies will greatly increase the data accuracy and provide much needed new insights into mechanisms and disease relevance. A good example of combining multiple technologies is a genome-wide siRNA screen searching for Wnt pathway regulators in a colorectal cancer cell line, which was conducted together with AP/MS studies of several proteins in this pathway [144]. Several relatively weak hits from AP/MS have been functionally confirmed by siRNA screening [144].

Moreover, the field of mathematical modeling is also growing fast. The new Markov network methodology has been proven effective in building neuron networks and incorporating different distributions such as Gaussian, Poisson, and Binomial in one model [145]. We hope that one day data produced by all different platforms will be integrated into one database and generate a super-interactive network for signaling pathways, which truly reflect the complex biological systems we are working on.

Acknowledgments

We thank Drs. David H. Hawke and Kathryn L. Hale for critical reading of the manuscript. X.L. is a recipient of a Computational Cancer Biology Training Program Fellowship supported by the Cancer Prevention and Research Institute of Texas. This work was supported in part with a startup fund from The University of Texas MD Anderson Cancer Center and Era of Hope Scholar Research Award W81XWH-09-1-0409 from the U.S. Department of Defense (J.C.). The University of Texas MD Anderson Cancer Center is supported by the National Institutes of Health Core Grant CA016672.

Abbreviations

TGF-β	transforming growth factor–beta
PPI	protein-protein interaction
PTM	post-translational modification
MS	mass spectrometry
MS/MS	tandem mass spectrometry
LC	liquid chromatography

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ICAT	isotope-coded affinity tags
SILAC	stable isotope labelling by amino acids in cell culture
iTRAQ	isobaric tags for relative and absolute quantification
TMTs	tandem mass tags
MS3	triple-stage mass spectrometry
Ub	ubiquitin
AP	affinity purification
TAP	tandem affinity purification
GO	gene ontology
co-IP	co-immunoprecipitation
FLAG	FLAG peptide tag
HA	human influenza hemagglutinin peptide tag
GFP	green fluorescent protein
TEV	TAP tags containing a tobacco etch virus protease cleavage site
EGF	epidermal growth factor
FDR	false discovery rate
NSAF	normalized spectral abundance factor
CompPASS	comparative proteomic analysis software suite
SAINT	significance analysis of interactomes
DUB	deubiquitinating enzyme
CRAPome	contaminant repository for affinity purification-mass spectrometry data
HCIP	high-confidence candidate interacting proteins
SFB	streptavidin-FLAG-S triple tags
BioID	proximity-dependent biotin identification
DPIM	the drosophila protein interaction mapping project

References

1. Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature. 2003; 425:577–84. [PubMed: 14534577]
2. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003; 422:198–207. [PubMed: 12634793]
3. Matthiesen R, Azevedo L, Amorim A, Carvalho AS. Discussion on common data analysis strategies used in MS-based proteomics. Proteomics. 2011; 11:604–19. [PubMed: 21241018]
4. Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001; 73:5683–90. [PubMed: 11774908]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

5. Domon B, Aebersold R. Mass spectrometry and protein analysis. Science. 2006; 312:212–7. [PubMed: 16614208]
6. 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; 17:994–9. [PubMed: 10504701]
7. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002; 1:376–86. [PubMed: 12118079]
8. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004; 3:1154–69. [PubMed: 15385600]
9. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, et al. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem. 2008; 80:2921–31. [PubMed: 18312001]
10. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem. 2003; 75:1895–904. [PubMed: 12713048]
11. Ting L, Rad R, Gygi SP, Haas W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods. 2011; 8:937–40. [PubMed: 21963607]
12. Choe L, D'Ascenzo M, Relkin NR, Pappin D, Ross P, Williamson B, et al. 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. Proteomics. 2007; 7:3651–60. [PubMed: 17880003]
13. Murphy JP, Everley RA, Coloff JL, Gygi SP. Combining Amine Metabolomics and Quantitative Proteomics of Cancer Cells Using Derivatization with Isobaric Tags. Anal Chem. 2014
14. Everley RA, Kunz RC, McAllister FE, Gygi SP. Increasing throughput in targeted proteomics assays: 54-plex quantitation in a single mass spectrometry run. Anal Chem. 2013; 85:5340–6. [PubMed: 23662842]
15. Altelaar AF, Frese CK, Preisinger C, Hennrich ML, Schram AW, Timmers HT, et al. Benchmarking stable isotope labeling based quantitative proteomics. J Proteomics. 2013; 88:14–26. [PubMed: 23085607]
16. Karp NA, Huber W, Sadowski PG, Charles PD, Hester SV, Lilley KS. Addressing accuracy and precision issues in iTRAQ quantitation. Mol Cell Proteomics. 2010; 9:1885–97. [PubMed: 20382981]
17. Shirran SL, Botting CH. A comparison of the accuracy of iTRAQ quantification by nLC-ESI MSMS and nLC-MALDI MSMS methods. J Proteomics. 2010; 73:1391–403. [PubMed: 20230925]
18. Kohlbacher O, Reinert K, Gropl C, Lange E, Pfeifer N, Schulz-Trieglaff O, et al. TOPP--the OpenMS proteomics pipeline. Bioinformatics. 2007; 23:e191–7. [PubMed: 17237091]
19. Rost HL, Schmitt U, Aebersold R, Malmstrom L. pyOpenMS: a Python-based interface to the OpenMS mass-spectrometry algorithm library. Proteomics. 2014; 14:74–7. [PubMed: 24420968]
20. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008; 26:1367–72. [PubMed: 19029910]
21. Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, et al. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc. 2009; 4:698–705. [PubMed: 19373234]
22. Matthiesen R. Virtual Expert Mass Spectrometrist v3.0: an integrated tool for proteome analysis. Methods Mol Biol. 2007; 367:121–38. [PubMed: 17185773]
23. Rodriguez-Suarez E, Gubb E, Alzueta IF, Falcon-Perez JM, Amorim A, Elortza F, et al. Virtual expert mass spectrometrist: iTRAQ tool for database-dependent search, quantitation and result storage. Proteomics. 2010; 10:1545–56. [PubMed: 20140907]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

24. de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature. 2008; 455:1251–4. [PubMed: 18820680]
25. Raso C, Cosentino C, Gaspari M, Malara N, Han X, McClatchy D, et al. Characterization of breast cancer interstitial fluids by TmT labeling, LTQ-Orbitrap Velos mass spectrometry, and pathway analysis. J Proteome Res. 2012; 11:3199–210. [PubMed: 22563702]
26. Dai L, He J, Liu Y, Byun J, Vivekanandan A, Pennathur S, et al. Dose-dependent proteomic analysis of glioblastoma cancer stem cells upon treatment with gamma-secretase inhibitor. Proteomics. 2011; 11:4529–40. [PubMed: 21932445]
27. Chen YS, Mathias RA, Mathivanan S, Kapp EA, Moritz RL, Zhu HJ, et al. Proteomics profiling of Madin-Darby canine kidney plasma membranes reveals Wnt-5a involvement during oncogenic H-Ras/TGF-beta-mediated epithelial-mesenchymal transition. Mol Cell Proteomics. 2011; 10:M110 001131.
28. Kim YM, Kim J, Heo SC, Shin SH, Do EK, Suh DS, et al. Proteomic identification of ADAM12 as a regulator for TGF-beta1-induced differentiation of human mesenchymal stem cells to smooth muscle cells. PLoS One. 2012; 7:e40820. [PubMed: 22808268]
29. Ali NA, McKay MJ, Molloy MP. Proteomics of Smad4 regulated transforming growth factor-beta signalling in colon cancer cells. Mol Biosyst. 2010; 6:2332–8. [PubMed: 20862427]
30. Macek B, Mann M, Olsen JV. Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol Toxicol. 2009; 49:199–221. [PubMed: 18834307]
31. Nita-Lazar A, Saito-Benz H, White FM. Quantitative phosphoproteomics by mass spectrometry: past, present, and future. Proteomics. 2008; 8:4433–43. [PubMed: 18846511]
32. Harsha HC, Pinto SM, Pandey A. Proteomic strategies to characterize signaling pathways. Methods Mol Biol. 2013; 1007:359–77. [PubMed: 23666735]
33. Kirkpatrick DS, Denison C, Gygi SP. Weighing in on ubiquitin: the expanding role of mass-spectrometry-based proteomics. Nat Cell Biol. 2005; 7:750–7. [PubMed: 16056266]
34. Shi Y, Xu P, Qin J. Ubiquitinated proteome: ready for global? Mol Cell Proteomics. 2011; 10:R110 006882. [PubMed: 21339389]
35. Mischerikow N, Heck AJ. Targeted large-scale analysis of protein acetylation. Proteomics. 2011; 11:571–89. [PubMed: 21246731]
36. Afjehi-Sadat L, Garcia BA. Comprehending dynamic protein methylation with mass spectrometry. Curr Opin Chem Biol. 2013; 17:12–9. [PubMed: 23333572]
37. Pan S, Chen R, Aebersold R, Brentnall TA. Mass spectrometry based glycoproteomics--from a proteomics perspective. Mol Cell Proteomics. 2011; 10:R110 003251. [PubMed: 20736408]
38. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER 3rd, Hurov KE, Luo J, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science. 2007; 316:1160–6. [PubMed: 17525332]
39. Olsen JV, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal. 2010; 3:ra3. [PubMed: 20068231]
40. Ali NA, Molloy MP. Quantitative phosphoproteomics of transforming growth factor-beta signaling in colon cancer cells. Proteomics. 2011; 11:3390–401. [PubMed: 21751366]
41. Tang LY, Deng N, Wang LS, Dai J, Wang ZL, Jiang XS, et al. Quantitative phosphoproteome profiling of Wnt3a-mediated signaling network: indicating the involvement of ribonucleoside-diphosphate reductase M2 subunit phosphorylation at residue serine 20 in canonical Wnt signal transduction. Mol Cell Proteomics. 2007; 6:1952–67. [PubMed: 17693683]
42. Kruger M, Kratchmarova I, Blagoev B, Tseng YH, Kahn CR, Mann M. Dissection of the insulin signaling pathway via quantitative phosphoproteomics. Proc Natl Acad Sci U S A. 2008; 105:2451–6. [PubMed: 18268350]
43. Amanchy R, Zhong J, Molina H, Chaerkady R, Iwahori A, Kalume DE, et al. Identification of c-Src tyrosine kinase substrates using mass spectrometry and peptide microarrays. J Proteome Res. 2008; 7:3900–10. [PubMed: 18698806]

Author Manuscript

44. Willert K, Brown JD, Danenberg E, Duncan AW, Weissman IL, Reya T, et al. Wnt proteins are lipid-modified and can act as stem cell growth factors. Nature. 2003; 423:448–52. [PubMed: 12717451]
45. Moren A, Hellman U, Inada Y, Imamura T, Heldin CH, Moustakas A. Differential ubiquitination defines the functional status of the tumor suppressor Smad4. J Biol Chem. 2003; 278:33571–82. [PubMed: 12794086]
46. Matsuura A, Ito M, Sakaidani Y, Kondo T, Murakami K, Furukawa K, et al. O-linked N-acetylglucosamine is present on the extracellular domain of notch receptors. J Biol Chem. 2008; 283:35486–95. [PubMed: 18948267]
47. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotechnol. 2005; 23:1391–8. [PubMed: 16273072]
48. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–9. [PubMed: 10802651]
49. Haw R, Hermjakob H, D'Eustachio P, Stein L. Reactome pathway analysis to enrich biological discovery in proteomics data sets. Proteomics. 2011; 11:3598–613. [PubMed: 21751369]
50. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013; 41:D808–15. [PubMed: 23203871]
51. Olsen JV, Mann M. Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol Cell Proteomics. 2013; 12:3444–52. [PubMed: 24187339]
52. Berggard T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. Proteomics. 2007; 7:2833–42. [PubMed: 17640003]
53. Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet. 2013; 14:35–48. [PubMed: 23207911]
54. Huen MS, Grant R, Manke I, Minn K, Yu X, Yaffe MB, et al. RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. Cell. 2007; 131:901–14. [PubMed: 18001825]
55. Liu T, Ghosal G, Yuan J, Chen J, Huang J. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. Science. 2010; 329:693–6. [PubMed: 20671156]
56. Yu X, Chini CC, He M, Mer G, Chen J. The BRCT domain is a phospho-protein binding domain. Science. 2003; 302:639–42. [PubMed: 14576433]
57. Lou Z, Minter-Dykhouse K, Wu X, Chen J. MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways. Nature. 2003; 421:957–61. [PubMed: 12607004]
58. Kim JE, Chen J, Lou Z. DBC1 is a negative regulator of SIRT1. Nature. 2008; 451:583–6. [PubMed: 18235501]
59. Huen MS, Sy SM, Chen J. BRCA1 and its toolbox for the maintenance of genome integrity. Nat Rev Mol Cell Biol. 2010; 11:138–48. [PubMed: 20029420]
60. Knuesel M, Wan Y, Xiao Z, Holinger E, Lowe N, Wang W, et al. Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. Mol Cell Proteomics. 2003; 2:1225–33. [PubMed: 12963786]
61. Luo Q, Nieves E, Kzhyshkowska J, Angeletti RH. Endogenous transforming growth factor-beta receptor-mediated Smad signaling complexes analyzed by mass spectrometry. Mol Cell Proteomics. 2006; 5:1245–60. [PubMed: 16582422]
62. Major MB, Camp ND, Berndt JD, Yi X, Goldenberg SJ, Hubbert C, et al. Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. Science. 2007; 316:1043–6. [PubMed: 17510365]
63. Brown KA, Ham AJ, Clark CN, Meller N, Law BK, Chytil A, et al. Identification of novel Smad2 and Smad3 associated proteins in response to TGF-beta1. J Cell Biochem. 2008; 105:596–611. [PubMed: 18729074]
64. Conrotto P, Yakymovych I, Yakymovych M, Souchelnytskyi S. Interactome of transforming growth factor-beta type I receptor (TbetaRI): inhibition of TGFbeta signaling by Epac1. J Proteome Res. 2007; 6:287–97. [PubMed: 17203972]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

65. Chaudhry SS, Cain SA, Morgan A, Dallas SL, Shuttleworth CA, Kielty CM. Fibrillin-1 regulates the bioavailability of TGFbeta1. J Cell Biol. 2007; 176:355–67. [PubMed: 17242066]

66. Angers S, Thorpe CJ, Biechele TL, Goldenberg SJ, Zheng N, MacCoss MJ, et al. The KLHL12-Cullin-3 ubiquitin ligase negatively regulates the Wnt-beta-catenin pathway by targeting Dishevelled for degradation. Nat Cell Biol. 2006; 8:348–57. [PubMed: 16547521]

67. Hilger M, Mann M. Triple SILAC to determine stimulus specific interactions in the Wnt pathway. J Proteome Res. 2012; 11:982–94. [PubMed: 22011079]

68. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002; 415:141–7. [PubMed: 11805826]

69. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature. 2002; 415:180–3. [PubMed: 11805837]

70. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature. 2006; 440:637–43. [PubMed: 16554755]

71. Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, et al. A protein complex network of Drosophila melanogaster. Cell. 2011; 147:690–703. [PubMed: 22036573]

72. Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, et al. Analysis of the human endogenous coregulator complexome. Cell. 2011; 145:787–99. [PubMed: 21620140]

73. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, et al. A protein interaction map of Drosophila melanogaster. Science. 2003; 302:1727–36. [PubMed: 14605208]

74. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, et al. A map of the interactome network of the metazoan C. elegans. Science. 2004; 303:540–3. [PubMed: 14704431]

75. Sardiu ME, Cai Y, Jin J, Swanson SK, Conaway RC, Conaway JW, et al. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. Proc Natl Acad Sci U S A. 2008; 105:1454–9. [PubMed: 18218781]

76. Sowa ME, Bennett EJ, Gygi SP, Harper JW. Defining the human deubiquitinating enzyme interaction landscape. Cell. 2009; 138:389–403. [PubMed: 19615732]

77. Joshi P, Greco TM, Guise AJ, Luo Y, Yu F, Nesvizhskii AI, et al. The functional interactome landscape of the human histone deacetylase family. Mol Syst Biol. 2013; 9:672. [PubMed: 23752268]

78. Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. Nature. 2010; 466:68–76. [PubMed: 20562859]

79. Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol. 2007; 8:645–54. [PubMed: 17593931]

80. Couzens AL, Knight JD, Kean MJ, Teo G, Weiss A, Dunham WH, et al. Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. Sci Signal. 2013; 6:rs15. [PubMed: 24255178]

81. Hauri S, Wepf A, van Drogen A, Varjosalo M, Tapon N, Aebersold R, et al. Interaction proteome of human Hippo signaling: modular control of the co-activator YAP1. Mol Syst Biol. 2013; 9:713. [PubMed: 24366813]

82. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol. 1999; 17:1030–2. [PubMed: 10504710]

83. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, et al. Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. Plant J. 2006; 46:1–13. [PubMed: 16553892]

84. Veraksa A, Bauer A, Artavanis-Tsakonas S. Analyzing protein complexes in Drosophila with tandem affinity purification-mass spectrometry. Dev Dyn. 2005; 232:827–34. [PubMed: 15704125]

85. Listgarten J, Emili A. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics. 2005; 4:419–34. [PubMed: 15741312]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

86. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004; 76:4193–201. [PubMed: 15253663]
87. Sarraf SA, Raman M, Guarani-Pereira V, Sowa ME, Huttlin EL, Gygi SP, et al. Landscape of the PARKIN-dependent ubiquitylome in response to mitochondrial depolarization. Nature. 2013; 496:372–6. [PubMed: 23503661]
88. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat Biotechnol. 2003; 21:315–8. [PubMed: 12577067]
89. Kratchmarova I, Blagoev B, Haack-Sorensen M, Kassem M, Mann M. Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. Science. 2005; 308:1472–7. [PubMed: 15933201]
90. Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods. 2007; 4:787–97. [PubMed: 17901868]
91. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007; 4:207–14. [PubMed: 17327847]
92. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. 2007; 4:923–5. [PubMed: 17952086]
93. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom. 1994; 5:976–89. [PubMed: 24226387]
94. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999; 20:3551–67. [PubMed: 10612281]
95. Fenyo D, Beavis RC. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal Chem. 2003; 75:768–74. [PubMed: 12622365]
96. Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, et al. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat Methods. 2011; 8:70–3. [PubMed: 21131968]
97. Foulds CE, Feng Q, Ding C, Bailey S, Hunsaker TL, Malovannaya A, et al. Proteomic analysis of coregulators bound to ERalpha on DNA and nucleosomes reveals coregulator dynamics. Mol Cell. 2013; 51:185–99. [PubMed: 23850489]
98. Christianson JC, Olzmann JA, Shaler TA, Sowa ME, Bennett EJ, Richter CM, et al. Defining human ERAD networks through an integrative mapping strategy. Nat Cell Biol. 2012; 14:93–105. [PubMed: 22119785]
99. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, et al. A global protein kinase and phosphatase interaction network in yeast. Science. 2010; 328:1043–6. [PubMed: 20489023]
100. Kwon Y, Vinayagam A, Sun X, Dephoure N, Gygi SP, Hong P, et al. The Hippo signaling pathway interactome. Science. 2013; 342:737–40. [PubMed: 24114784]
101. Wang W, Li X, Huang J, Feng L, Dolinta KG, Chen J. Defining the protein-protein interaction network of the human hippo pathway. Mol Cell Proteomics. 2014; 13:119–31. [PubMed: 24126142]
102. Choi H, Glatter T, Gstaiger M, Nesvizhskii AI. SAINT-MS1: protein-protein interaction scoring using label-free intensity data in affinity purification-mass spectrometry experiments. J Proteome Res. 2012; 11:2619–24. [PubMed: 22352807]
103. Teo G, Liu G, Zhang J, Nesvizhskii AI, Gingras AC, Choi H. SAINTexpress: Improvements and additional features in Significance Analysis of Interactome software. J Proteomics. 2013
104. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods. 2013; 10:730–6. [PubMed: 23921808]
105. Varjosalo M, Sacco R, Stukalov A, van Drogen A, Planyavsky M, Hauri S, et al. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. Nat Methods. 2013; 10:307–14. [PubMed: 23455922]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

106. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, et al. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007; 2:2366–82. [PubMed: 17947979]
107. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504. [PubMed: 14597658]
108. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, et al. A travel guide to Cytoscape plugins. Nat Methods. 2012; 9:1069–76. [PubMed: 23132118]
109. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5:R80. [PubMed: 15461798]
110. Yan Y, Marriott G. Analysis of protein interactions using fluorescence technologies. Curr Opin Chem Biol. 2003; 7:635–40. [PubMed: 14580569]
111. Piehler J. New methodologies for measuring protein interactions in vivo and in vitro. Curr Opin Struct Biol. 2005; 15:4–14. [PubMed: 15718127]
112. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989; 340:245–6. [PubMed: 2547163]
113. Endoh H, Walhout AJ, Vidal M. A green fluorescent protein-based reverse two-hybrid system: application to the characterization of large numbers of potential protein-protein interactions. Methods Enzymol. 2000; 328:74–88. [PubMed: 11075339]
114. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003; 31:258–61. [PubMed: 12519996]
115. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006; 34:D535–9. [PubMed: 16381927]
116. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, et al. IntAct--open source resource for molecular interaction data. Nucleic Acids Res. 2007; 35:D561–5. [PubMed: 17145710]
117. Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 2003; 31:248–50. [PubMed: 12519993]
118. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the database of interacting proteins. Nucleic Acids Res. 2000; 28:289–91. [PubMed: 10592249]
119. Prasad TS, Kandasamy K, Pandey A. Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. Methods Mol Biol. 2009; 577:67–79. [PubMed: 19718509]
120. Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, Venugopal AK, et al. NetPath: a public resource of curated signal transduction pathways. Genome Biol. 2010; 11:R3. [PubMed: 20067622]
121. Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol. 2004; 6:97–105. [PubMed: 14743216]
122. Glatter T, Schittenhelm RB, Rinner O, Roguska K, Wepf A, Junger MA, et al. Modularity and hormone sensitivity of the Drosophila melanogaster insulin receptor/target of rapamycin interaction proteome. Mol Syst Biol. 2011; 7:547. [PubMed: 22068330]
123. Justice RW, Zilian O, Woods DF, Noll M, Bryant PJ. The Drosophila tumor suppressor gene warts encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. Genes Dev. 1995; 9:534–46. [PubMed: 7698644]
124. Xu T, Wang W, Zhang S, Stewart RA, Yu W. Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase. Development. 1995; 121:1053–63. [PubMed: 7743921]
125. Kango-Singh M, Nolo R, Tao C, Verstreken P, Hiesinger PR, Bellen HJ, et al. Shar-pei mediates cell proliferation arrest during imaginal disc growth in Drosophila. Development. 2002; 129:5719–30. [PubMed: 12421711]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

126. Tapon N, Harvey KF, Bell DW, Wahrer DC, Schiripo TA, Haber D, et al. salvador Promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines. Cell. 2002; 110:467–78. [PubMed: 12202036]
127. Harvey KF, Pfleger CM, Hariharan IK. The Drosophila Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. Cell. 2003; 114:457–67. [PubMed: 12941274]
128. Wu S, Huang J, Dong J, Pan D. hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts. Cell. 2003; 114:445–56. [PubMed: 12941273]
129. Lai ZC, Wei X, Shimizu T, Ramos E, Rohrbaugh M, Nikolaidis N, et al. Control of cell proliferation and apoptosis by mob as tumor suppressor, mats. Cell. 2005; 120:675–85. [PubMed: 15766530]
130. Huang J, Wu S, Barrera J, Matthews K, Pan D. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila Homolog of YAP. Cell. 2005; 122:421–34. [PubMed: 16096061]
131. Hamaratoglu F, Willecke M, Kango-Singh M, Nolo R, Hyun E, Tao C, et al. The tumour-suppressor genes NF2/Merlin and Expanded act through Hippo signalling to regulate cell proliferation and apoptosis. Nat Cell Biol. 2006; 8:27–36. [PubMed: 16341207]
132. Pan D. Hippo signaling in organ size control. Genes Dev. 2007; 21:886–97. [PubMed: 17437995]
133. Zhao B, Lei QY, Guan KL. The Hippo-YAP pathway: new connections between regulation of organ size and cancer. Curr Opin Cell Biol. 2008; 20:638–46. [PubMed: 18955139]
134. Harvey KF, Zhang X, Thomas DM. The Hippo pathway and human cancer. Nat Rev Cancer. 2013; 13:246–57. [PubMed: 23467301]
135. Friedman AA, Tucker G, Singh R, Yan D, Vinayagam A, Hu Y, et al. Proteomic and functional genomic landscape of receptor tyrosine kinase and ras to extracellular signal-regulated kinase signaling. Sci Signal. 2011; 4:rs10. [PubMed: 22028469]
136. Kyriakakis P, Tipping M, Abed L, Veraksa A. Tandem affinity purification in Drosophila: the advantages of the GS-TAP system. Fly (Austin). 2008; 2:229–35. [PubMed: 18719405]
137. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol. 2012; 196:801–10. [PubMed: 22412018]
138. Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN, Mohan SS, et al. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet. 2006; 38:285–93. [PubMed: 16501559]
139. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, et al. Comparative assessment of large-scale data sets of protein-protein interactions. Nature. 2002; 417:399–403. [PubMed: 12000970]
140. Moya IM, Halder G. Discovering the Hippo pathway protein-protein interactome. Cell Res. 2014; 24:137–8. [PubMed: 24418760]
141. Figeys D, McBroom LD, Moran MF. Mass spectrometry for the study of protein-protein interactions. Methods. 2001; 24:230–9. [PubMed: 11403572]
142. McHugh L, Arthur JW. Computational methods for protein identification from mass spectrometry data. PLoS Comput Biol. 2008; 4:e12. [PubMed: 18463710]
143. Bell AW, Deutsch EW, Au CE, Kearney RE, Beavis R, Sechi S, et al. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. Nat Methods. 2009; 6:423–30. [PubMed: 19448641]
144. Major MB, Moon RT. “Omic” risk assessment. Sci Signal. 2009; 2:eg7. [PubMed: 19471021]
145. Tomson SN, Narayan M, Allen GI, Eagleman DM. Neural networks of colored sequence synesthesia. J Neurosci. 2013; 33:14098–106. [PubMed: 23986245]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

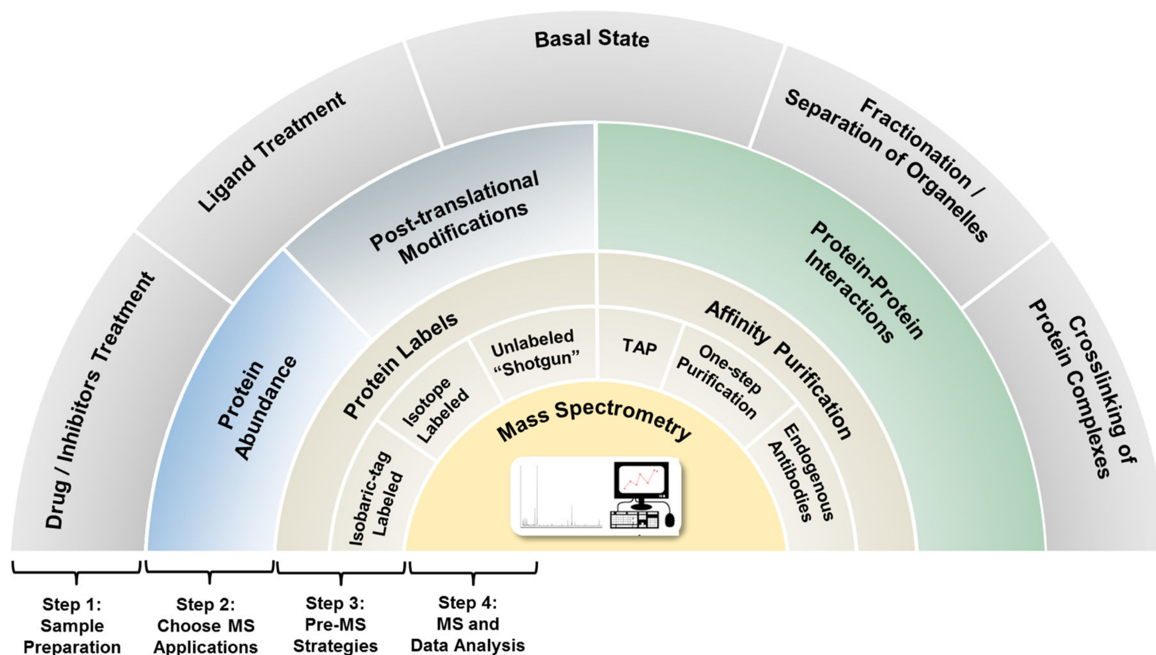

Figure 1. An integrated view of the MS application and strategies in signaling pathway studies

Step 1: Sample preparation. Basal state cells could be used directly for MS studies, or treated with ligand, drug, or inhibitors to obtain different states in signaling events. Cells may be subjected to fractionation or other approaches to separate certain fractions/ organelles, or cross-linked with chemicals, to monitor the proteome/interactome changes of a part of the cells. Step 2: Choose MS applications. MS could be used to monitor the overall protein abundance, post-translational modifications, and protein-protein interactions. Step 3: Pre-MS strategies. Protein complexes could directly subjected to MS ("shotgun" methods) or isotope/isobaric labeling prior to MS. For affinity purifications, antibodies against endogenous proteins could be used to bring down the endogenous protein complexes. One-step or tandem affinity purifications (TAP) could be used to bring down epitope-tagged bait proteins. Step 4: MS and data analysis. Different MS equipment could be used to generate peptide readouts. Processed MS/MS spectra are searched against protein sequence databases thereby assigning MS/MS spectra to peptides, which can be subsequently mapped to proteins. Additional algorithms and software are commonly used in the follow-up data analysis to generate biologically relevant readouts.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

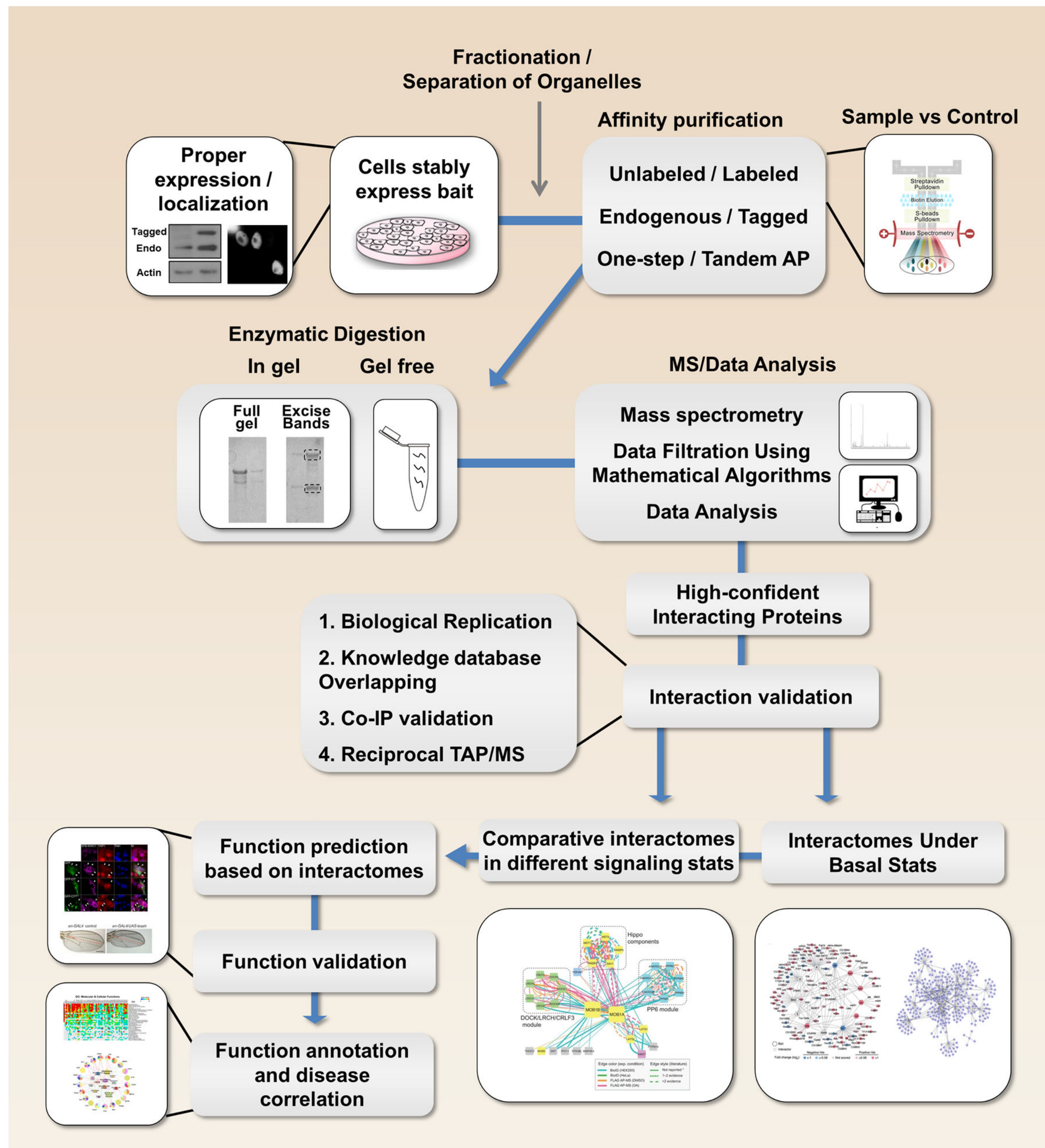

Figure 2. An example of current workflow in AP/MS studies of a signaling pathway

Bait proteins are usually epitope tagged and stably expressed in the cell of interest. Western blot analysis and immunofluorescence are usually carried out to ensure the proper localization and expression level of bait proteins. Ideally, the expression should be similar to the endogenous level of the protein. Cell lysate could run on sodium dodecyl sulfate–polyacrylamide gel electrophoresis and digested by enzymatic reaction (i.e., trypsin) to short peptides. Samples could be labeled by isotope/isobaric tags before the one-step or tandem affinity purification. Control purifications of mock or unrelated proteins are usually performed with the same strategies to eliminate the common contaminants and abundant proteins in AP. MS is usually performed with biological and technical replicates, and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

followed by data analysis, to generate a list of high-confidence interactions, which could be used to build up the interactomes. But to ensure the quality of the dataset, data validation including knowledge database overlapping, co-IP, and reciprocal AP/MS validation of a subset of the interactomes, is needed. In-depth functional validation could be conducted for several proteins of interest to prove the dataset can lead to biological meaningful results. The interactomes could be used to predict bait and prey functions and their diseases correlations. The example snapshots were taken from Couzens *et al*., Kwon *et al*., Wang *et al*., and Li *et al*.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

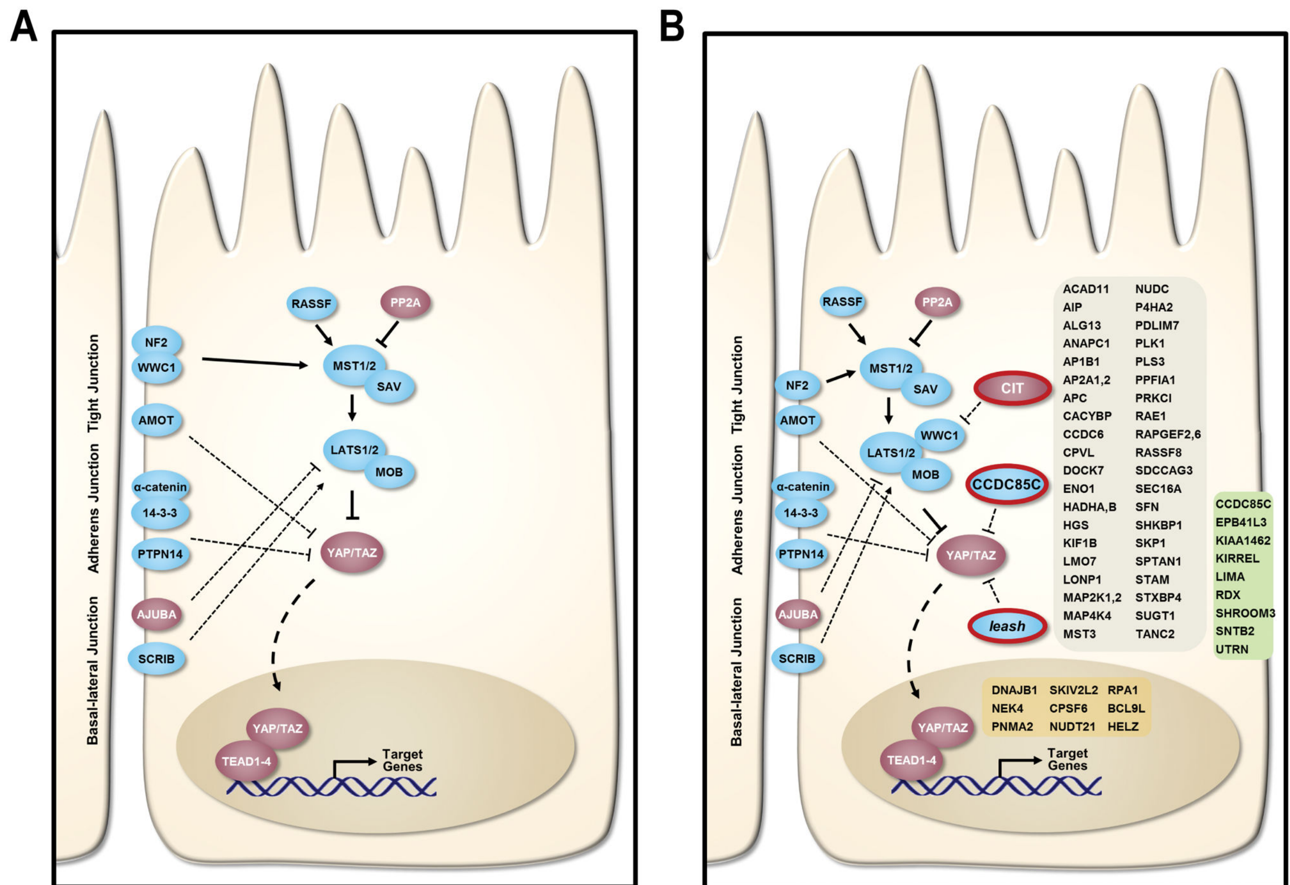

Figure 3. An overview of the human Hippo signaling pathway before and after proteomics studies

(A) The core knowledge of the Hippo signaling pathway before proteomics studies. Red bubbles indicate positive regulators and blue bubbles indicate negative regulators of Hippo signaling. (B) The Hippo signaling pathway refined with newly identified components. The three experimentally validated regulators CIT, CCDC85C, and *leash* are highlighted. Other novel HCIPs shown in at least two datasets are shown with different subcellular localizations (green: on membrane; tan: nuclear; grey: cytoplasmic and others). The list was generated by Moya *et al*. using the datasets from Couzens *et al*., Kwon *et al*., and Wang *et al*.