

For reprint orders, please contact reprints@expert-reviews.com

Protein complexes: the forest and the trees

Expert Rev. Proteomics 6(1), 5–10 (2009)



Guillaume Adelmant

Department of Cancer Biology and Blais Proteomics Center, Dana-Farber Cancer Institute, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA



Jarrod A Marto

*Author for correspondence
Department of Cancer Biology, Dana-Farber Cancer Institute, 44 Binney Street, Smith 1158A, Boston, MA 02115-6084, USA
Tel.: +1 617 632 3150
Fax: +1 617 582 7737
jarrod_marto@dfci.harvard.edu*

“...when viewed in the context of the extraordinary complexity presented by biological systems, we see great opportunity for significant improvements in the various techniques that constitute the systems biology toolbox.”

There is growing appreciation that control of cellular processes occurs through a delicate balance of protein post-translational modifications and interactions with other proteins, DNA, RNA and small-molecule metabolites. Similarly, phenotypic effects arising from specific genetic variations often manifest through a complex and dynamic network of interacting biomolecules. These observations have catalyzed efforts to move beyond the study of individual molecules towards the analysis of pathways and networks in order to understand cellular response to stress, infection, disease and other perturbations at the ‘systems’ level. Beginning in the 1980s, large-scale DNA sequencing launched genomics as a research field and, since then, an ever-expanding cottage industry of ‘-omics’ disciplines has appeared on the horizon. Ideally, a combination of these disparate data will generate a complete bioparts list with sufficient temporal and spatial resolution to support predictions of phenotype as a function of biological perturbation. Unfortunately, inspection of current -omics technologies reveals that they are not well matched in terms of overall throughput, data quality, standardization and other figures of merit. Indeed these methods span the full spectrum of technological maturity; gene chips and microarrays appear nearly commoditized as compared with proteomics and metabolomics. Similarly, computational approaches for integration of these data are at a primitive state of development. Interestingly, many of the limitations of -omics technologies, noted in a perspective published in 2000, are still pertinent today. Nonetheless, when viewed in the context of the extraordinary complexity presented by biological systems, we see

great opportunity for significant improvements in the various techniques that constitute the systems biology toolbox. This brief report will provide an overview of the challenges inherent to analysis of multicomponent protein complexes.

In practice yeast-two-hybrid (Y2H) and affinity purification (AP) followed by mass spectrometry (MS) are the two most widely used techniques to study protein–protein interactions [1,2]. Not surprisingly, these approaches provide complimentary information and analytical figures of merit; the choice of which technique to use should be driven by appropriate considerations of experimental context and the biological question at hand. For example, Y2H is amenable to high-throughput automation and is well suited for systematic identification of binary protein–protein interactions. However, because the measurements are made under nonphysiologic conditions, Y2H-derived interactions are often subjected to multiple levels of biochemical verification. Conversely, AP-MS detects protein complexes that assemble within a native environment, and provides a ready means to monitor post-translational modification states and relative expression levels as a function of biological perturbation. However, the state of automation generally lags behind that of Y2H and hence fewer examples of system-wide studies are available for AP-MS [3–8]. As with many -omics technologies, the data from Y2H and AP-MS are subject to high error rates if experiments are not performed with appropriate controls. By way of comparison, consider a forest viewed from different vantage points (FIGURE 1). From above (Y2H), one sees an unbroken canopy (network) composed of many intertwined branches (interactions); a dense canopy makes it

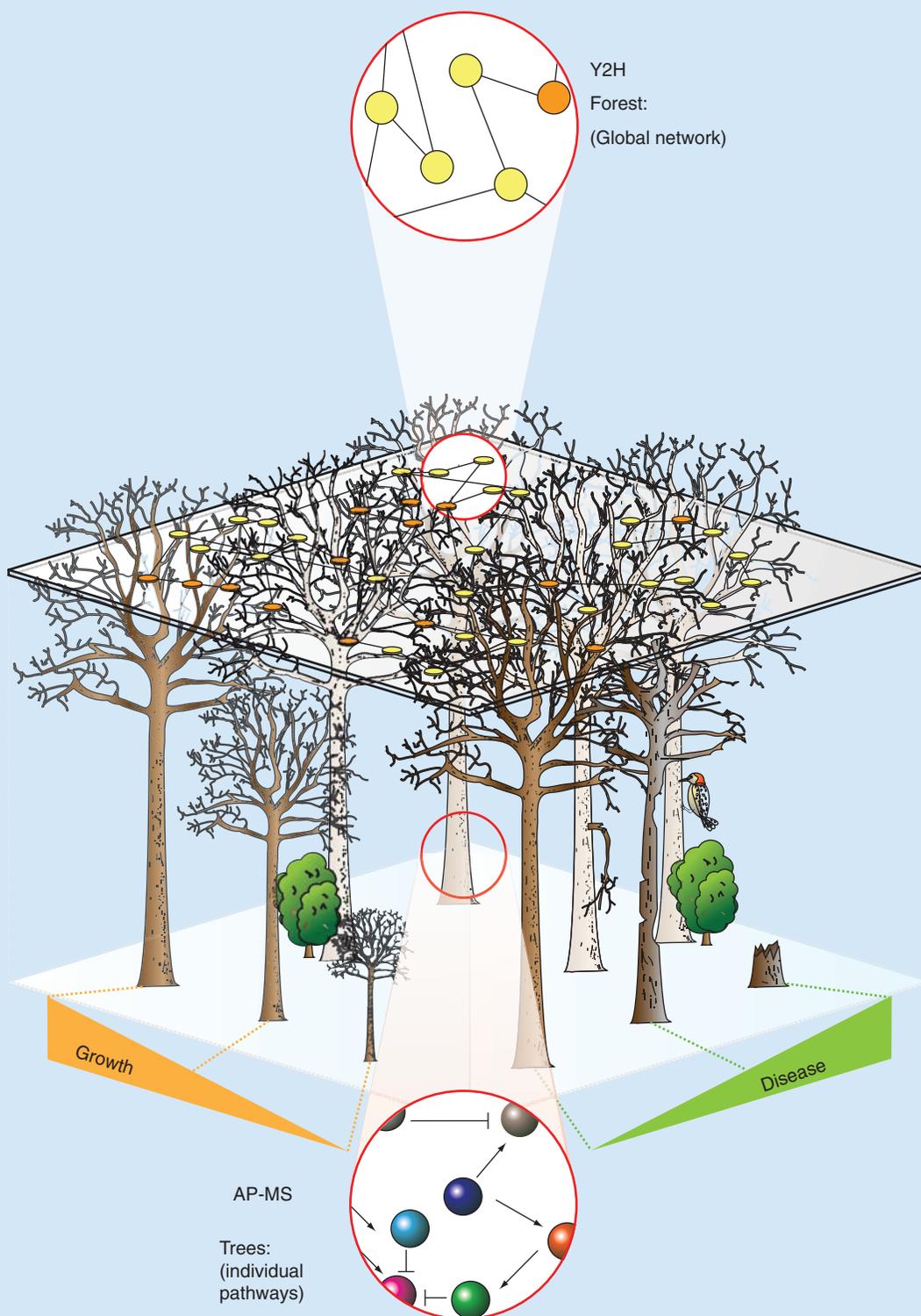


Figure 1. The forest (Y2H) and the trees (AP-MS). High-throughput Y2H measurements provide data akin to an aerial view of a forest. A dense canopy obscures information on individual trees. Ground-level data from AP-MS provides a detailed view of a few trees, but is less amenable to systems-level measurements.
 AP-MS: Affinity purification followed by mass spectrometry; Y2H: Yeast-two-hybrid.

difficult to discern individual trees (pathways). In fact, with increased altitude (throughput) one can view (detect) the entire canopy, although details of the forest floor become completely obscured (nonphysiological conditions). Conversely, individual trees (pathways) are readily identified from the ground (AP-MS). However, from this perspective it is virtually impossible to view (detect) the entire forest (network) without walking from tree to tree (low throughput). Standing in close proximity to one, or a few, trees facilitates detailed inspection of their biological condition (quantitative proteomics). Nearby foliage and inhabitants can interfere and complicate the analysis (nonspecific contamination). Y2H has been the subject of several recent reviews [9,10]. We will focus for the remainder of this article on the use of AP-MS for the analysis of protein complexes. In particular, we will discuss details pertinent to practitioners, especially those parameters that impact enrichment yield and specificity.

What is a protein complex?

It is instructive to consider two working definitions of 'protein complex.' Since most cellular functions are carried out by proteins carefully assembled in supramolecular structures [11], we might define protein complexes as the functional units of the proteome (e.g., RNA polymerase II, actin filaments and spliceosome). From a biochemical perspective a protein complex is typically described as the association of two or more polypeptides that survives some type of purification. Over the years, numerous methods have been employed for the purification of protein complexes. These may be classified by their mechanism of action: chemical (chromatography), biophysical (density, isoelectric point) and affinity interactions (immunological or other). Chromatography-based methods have a long history in the purification of proteins and associated complexes; however, they generally require a high level of expertise and extensive optimization for each protein of interest. Fractionation based on density has been used for the purification of very large protein complexes (e.g., ribosomes [12]) but provides limited resolution; and, as such, this technique is typically used as a discrete step within a more complex purification strategy. The purification of protein complexes by immunological methods has been used extensively over the years but is contingent on the availability of antibodies that are specific for the target of interest and provide acceptable yield. While most antibodies do not meet both criteria, biomedical researchers often prefer immunoaffinity-based methods due to widespread familiarity with the required reagents and modest equipment costs. Epitope-tagging strategies, whereby small peptide sequences recognized by well-characterized antibodies are added to a protein of interest, have been developed [13]. This general tagging strategy now includes affinity reagents that are not based solely on antibody-antigen interaction. Recognizing that multistage purification would provide higher specificity, Rigaut *et al.* developed a two-step purification method based on the use of sequential tags. In a variation on this theme, Nakatani *et al.* introduced the use of sequential immunoaffinity purification directed against two different peptide tags [14,15]. The aforementioned approaches are now generally referred to as tandem affinity purification (TAP).

Many different tags are now available: Flag epitope, HA, V5, GST, Protein A, CBP and StrepII, among others [16]. The choice should be guided by their efficacy. In general, smaller tags are less likely to change the structure or localization, or generally interfere with protein function. However, as a precaution, potential interference should be evaluated carefully (by immunofluorescence or cell fractionation for localization, or by any functional assay, if available) before going any further with the purification and downstream analysis. In addition, it is worth noting that the location of the tag (N-terminal or C-terminal) can also impact a protein's function or disrupt protein-protein interactions. Often, a small-scale purification using the two versions of the tagged protein can help to evaluate this possibility.

Specificity

The greatest advantage of the TAP protocol is that it results in a very high enrichment factor under nearly physiological conditions. Consider that during the procedure, both the capture and the elution contribute to the specificity of the purification. Hence a TAP procedure encompasses four discrete steps that collectively provide very high specificity to the purification process. In some cases, however, the result of a tandem affinity purification may actually correspond to a mixture of several discrete complexes, originating for example from distinct subcellular compartments. In such cases, additional purification steps, such as subcellular fractionation prior to TAP, may be beneficial.

Contamination

Despite the remarkable specificity of the TAP procedure, several contaminants have been observed repeatedly in independent experiments across multiple laboratory. These contaminants are typically abundant cellular proteins that have low affinity for the matrix on which the affinity reagents are immobilized, constituting what has recently been described as the 'bead proteome' [17]. Chemical contamination can also be problematic. In some purification procedures, for example, the last step consists of the competitive elution of the protein complex by a large excess of the peptide epitope. This peptide can dramatically interfere with subsequent MS analysis of the digested protein complex and, therefore, must be quantitatively removed. Similarly, detergents commonly used to minimize nonspecific interactions may have a detrimental effect at many stages of the liquid chromatography-MS/MS analysis. In part because of their abundance and molecular heterogeneity, these contaminants are particularly difficult to eliminate. As a result, practitioners must be vigilant in monitoring this source of contamination.

Dynamic measurements

One of the major advantages of AP-MS is that the protein complex of interest is purified from a physiologically relevant context. For this reason, one can study potential changes in complex composition in response to a variety of intra- and extracellular stimuli. Relative changes in composition (using the tagged protein as a reference) or changes in post-translational modification status can be measured easily by quantitative MS using

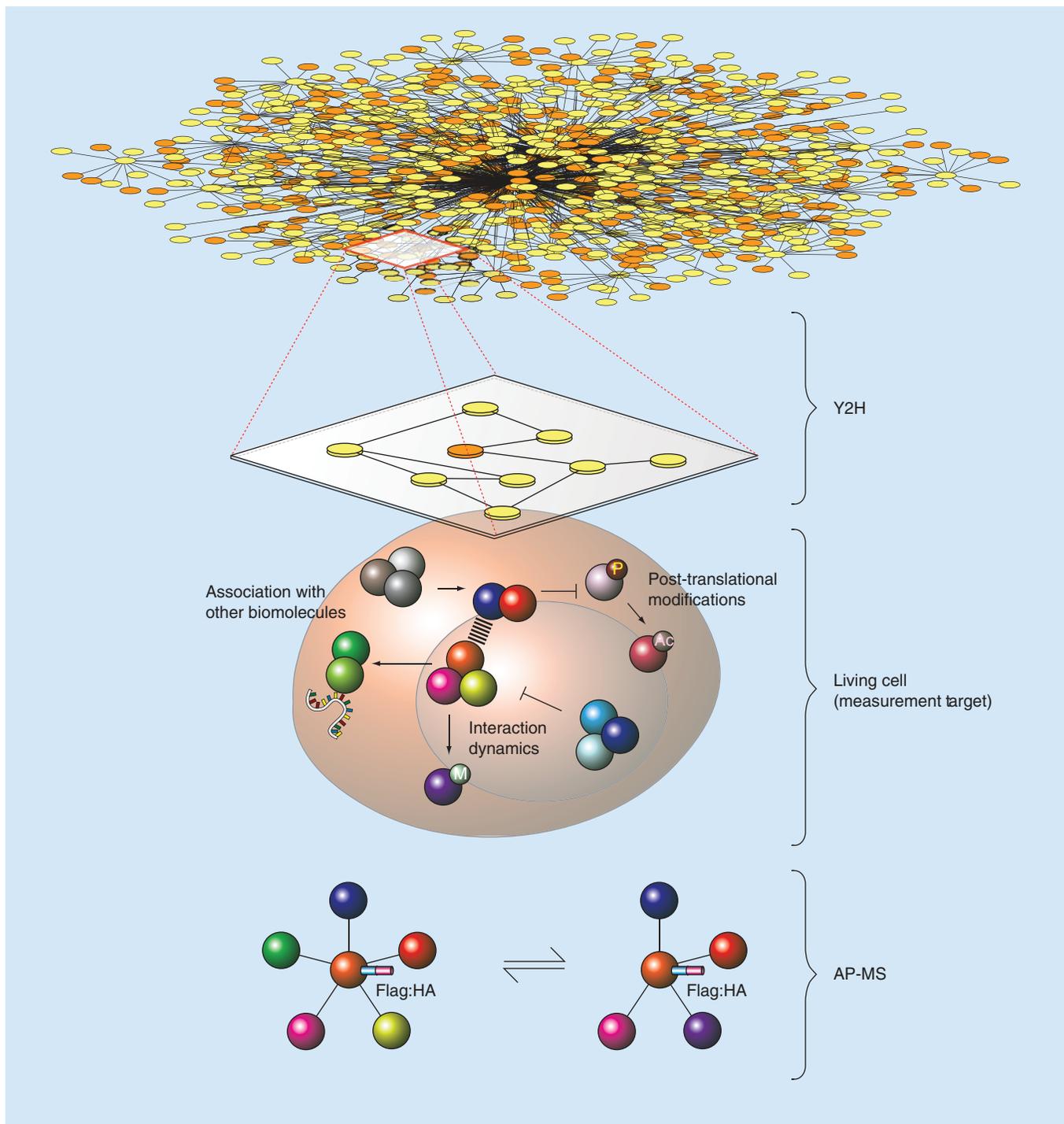


Figure 2. Protein–protein interactions, Y2H and AP-MS. Proteome-scale Y2H representation of cellular protein interactions (top). In this view, each edge represents a binary protein–protein interaction. Y2H interactions do not translate to all cell states and do not account for post-translational modifications (middle). AP-MS provides for the detection of protein complexes assembled under physiological conditions (bottom). Although more detailed in terms of dynamics and modifications, relative throughput is low for AP-MS and low-affinity interactions may be missed. AP-MS: Affinity purification followed by mass spectrometry; Y2H: Yeast-two-hybrid.

stable isotopic tags or so-called label-free methods [18]. These approaches work especially well for measuring changes affecting strong interactions between the target protein and its partners. Interactors with high on/off rates may be lost during purification,

complicating AP-MS-based analyses. Recently, new strategies have been introduced to improve yield for low-affinity protein partners and facilitate the identification of specific versus nonspecific interactions [19,20].

Other considerations

The choice of the cell type from which to purify a given protein should result from the evaluation of practical and biological considerations. The information gained from the purification of a transcription factor expressed in neurons will probably be richer if the TAP experiment is carried out in neurons or neuron-like cells. However, the large number of tissue culture plates necessary for a usable purification may be prohibitive. In addition to allowing purification from yeast or mammalian cells, the TAP procedure has also been used to purify protein complexes from mice, rice and other complex organisms [21–23].

Like endogenous proteins, the expression level for tagged proteins varies over a wide range, from very low for some transcription factors to very high for some structural proteins. As a consequence, it is difficult to predict how much starting material (or how many cells) will be necessary for the AP-MS analysis of a particular protein complex. In addition, some interactors in the purified complex may be present at substoichiometric levels, hence reproducible identification may require more material. This is particularly true for transient interactors with high dissociation constants.

In theory, the TAP procedure could be automated easily using either a liquid pipetting platform or a magnetic particle processor. A particular challenge facing automation will be establishing the best possible compromise between throughput, yield and specificity.

Expert commentary

In summary, the nature and quality of data that arise from protein–protein interaction studies depends on the experimental question at hand. The information derived from Y2H provides a large framework that represents potential connections between

proteins, and is a valuable resource to map cellular networks at a system level. While copious, high-throughput, binary protein interaction data have been generated, there is growing debate as to whether significant biological insight can be derived directly from large-scale, static networks. Recently this paradigm has been extended to so-called disease networks in an effort to correlate phenotype with protein interaction network topology [24]. However, high-throughput protein interaction studies do not typically account for post-translational modifications, often making the leap to functional insight difficult at best. AP-MS is well suited to the study of discrete complexes, rather than extended network topology, and provides an opportunity to monitor the dynamics of individual complexes members (FIGURE 2). While implementation at the whole-proteome scale encompasses many technical challenges, AP-MS will pay dividends along the way, through the production of quantitative data that describe individual protein complexes and provide a direct path for validation and functional verification using traditional biochemical-based assays.

Acknowledgement

The authors thank Eric Smith for valuable discussion and preparation of the figures.

Financial & competing interests disclosure

This work was supported by the Dana-Farber Cancer Institute and the National Human Genome Research Institute (P50HG004233). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

- Kohn KW, Aladjem MI, Weinstein JN, Pommier Y. Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. *Mol. Biol. Cell* 17(1), 1–13 (2006).
- Brent R. Genomic biology. *Cell* 100(1), 169–183 (2000).
- Arifuzzaman M, Maeda M, Itoh A *et al.* Large-scale identification of protein–protein interaction of *Escherichia coli* K-12. *Genome Res.* 16(5), 686–691 (2006).
- Ewing RM, Chu P, Elisma F *et al.* Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Syst. Biol.* 3, 89 (2007).
- Gavin AC, Aloy P, Grandi P *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440(7084), 631–636 (2006).
- Gavin AC, Bosche M, Krause R *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415(6868), 141–147 (2002).
- Ho Y, Gruhler A, Heilbut A *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415(6868), 180–183 (2002).
- Krogan NJ, Cagney G, Yu H *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440(7084), 637–643 (2006).
- Cusick ME, Klitgord N, Vidal M, Hill DE. Interactome: gateway into systems biology. *Hum. Mol. Genet.* 14(Spec. 2), R171–R181 (2005).
- Suter B, Kittanakom S, Stagljar I. Two-hybrid technologies in proteomics research. *Curr. Opin. Biotech.* 19(4), 316–323 (2008).
- Alberts B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92(3), 291–294 (1998).
- Merrick WC. Purification of protein synthesis initiation factors from rabbit reticulocytes. *Methods Enzymol.* 60, 101–108 (1979).
- Jarvik JW, Telmer CA. Epitope tagging. *Ann. Rev. Genet.* 32, 601–618 (1998).
- Nakatani Y, Ogrzyzko V. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol.* 370, 430–444 (2003).
- 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.* 17(10), 1030–1032 (1999).
- Collins MO, Choudhary JS. Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr. Opin. Biotechnol.* 19(4), 324–330 (2008).

- 17 Trinkle-Mulcahy L, Boulon S, Lam YW *et al.* Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell. Biol.* 183(2), 223–239 (2008).
- 18 Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 422(6928), 198–207 (2003).
- 19 Mousson F, Kolkman A, Pijnappel WW, Timmers HT, Heck AJ. Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. *Mol. Cell. Proteomics* 7(5), 845–852 (2008).
- 20 Wang X, Huang L. Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Mol. Cell. Proteomics* 7(1), 46–57 (2008).
- 21 Kyriakakis P, Tipping M, Abed L, Veraksa A. Tandem affinity purification in *Drosophila*: the advantages of the GS-TAP system. *Fly* (2008) (Epub ahead of print).
- 22 Rohila JS, Chen M, Chen S *et al.* Protein–protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J.* 46(1), 1–13 (2006).
- 23 Testa G, Zhang Y, Vintersten K *et al.* Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nature Biotech.* 21(4), 443–447 (2003).
- 24 Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc. Natl Acad. Sci. USA* 104(21), 8685–8690 (2007).