hospital in Rotterdam, the Netherlands, and lead author of one of the studies.

In his protocol, Poot pulled interacting proteins from cells using nuclear extracts expressing different Flag-tagged transcription factors. He added a nuclease to his reactions to remove DNA and eliminate possible artefacts caused by proteins binding to it. "Transcription factors bind to DNA so you are likely to pull out DNA-binding factors that are not directly interacting," he explains. Purifying many different transcription factors with the same protocol also enabled the researchers to determine which interactions were most likely to be specific. For example, proteins that consistently co-purified with all transcription factors would be treated as unlikely to indicate a genuine interaction.

Calling out false positives — reported interactions that don't actually occur — and false negatives — interactions that do occur but are not picked up by the experimental protocol or are discarded — is one of the main challenges in the field. "Normally when you do a coIP followed by MS you will get hundreds of protein candidates interacting with any one bait," says Wade Harper, a cell biologist at Harvard Medical School in Boston, Massachusetts. "When you weed out all the stochastic and non-specific interactions you end up with many fewer proteins. Some proteins in large complexes might have 30–50 partners, others only 4–5."

One way in which researchers increase the accuracy of their results is to use more than one method (for example, Y2H plus LUMIER) to detect the interactions. But the definition of a 'real' interaction depends on the context. "Does a real interaction mean that two proteins interact if they are placed next to each other in a test tube, or that they must interact in a cell? Or does real mean that the interaction should have a biological function?" asks Ideker. Researchers can home in on functional interactions by combining data on interactions with other types of biological information, such as genetic interactions, protein localizations or gene expression. For instance, proteins whose genes are co-expressed are likely to interact with each other or to be part of the same complex or pathway.

Many tools are available on the web for integrating different types of information about a given protein or gene. One is GeneMANIA, developed by Bader's group in collaboration with Quaid Morris, a computational biologist also at the University of Toronto. A user enters the gene names into GeneMANIA; the program provides a list of genes that are functionally similar or have shared properties, such as similar expression or localization, and then displays a proposed interaction network, showing relationships among the genes and the type of data used to gather that information. The user can click on any node to obtain information about the gene and on any link to obtain information about their relationship (such as citations for any published studies or other sources of data). "It's like a Google for genetic and protein information," says Bader.

Other web-based interfaces that predict gene functions include STRING (http://string-db.org) developed at the European Molecular Biology Laboratory in Heidelberg, Germany. It hunts for protein interactions on the basis of genomic context, high-throughput experiments, co-expression and data from the literature.

KEEPING SCORE

To select real protein–protein interactions, Harper and some members of his lab, Matt Sowa and Eric Bennett, developed a software platform called CompPASS to assign confidence scores to an interaction detected by MS[5]. CompPASS takes data sets of interacting proteins (including those identified in experiments) and measures frequency, abundance and reproducibility of interactions to calculate the score.

This year, Harper used CompPASS to identify interactions among proteins involved in autophagy, the process by which cellular proteins and organelles are engulfed into vesicles and delivered to the lysosome to be degraded. Starting with 32 proteins known to have a role in autophagy, they identified 2,553 interacting proteins using coIP–MS. CompPASS then narrowed the list down to 409 high-confidence interacting proteins with 751 interactions[6].

Ideker's group used a different approach to map interactions among human mitogen-activated protein kinases (MAPKs), which respond to external stimuli and regulate cell function. Having used Y2H to identify more

Tools for the search

HYBRIGENICS

Methods such as the yeast two-hybrid system allow scientists to work out which proteins interact.

The two main methods for finding protein–protein interactions are the yeast two-hybrid (Y2H) system and co-immunoprecipitation followed by mass spectrometry. Several companies sell reagents for both approaches. Invitrogen of Carlsbad, California, sells the ProQuest Two-Hybrid System with Gateway Technology. This is based on Y2H, with modifications to decrease false-positive results and allow rapid characterization, says the company. Other firms provide vectors used to produce proteins with affinity tags, which can easily be immunoprecipitated along with other interacting proteins. A polypeptide tag called Flag is popular among researchers, and Sigma Aldrich of St Louis, Missouri, provides several Flag-genes for purchase. Promega in Madison, Wisconsin, has the HaloTag technology, in which a protein of interest is expressed in fusion with a tag protein engineered from a bacterial enzyme. This tag can be used to purify the protein, and any interacting with it, by binding to a resin. The tag is cleaved off using a protease.

For researchers who don't have the time or infrastructure to do the experiments, companies such as Hybrigenics in Paris and Dualsystems Biotech of Schlieren, Switzerland, offer Y2H-based screening. "We have complex libraries with ten times more independent clones than most other libraries, which we screen to saturation. And rather than screening full-length proteins, we screen for interactions with domains," says Etienne Formstecher, director of scientific projects and sales at Hybrigenics. "Full-length proteins can have some domains buried and not available to interact, at least in yeast where you may not have signals to unlock a closed protein conformation." A customer is given a list of proteins that interact with the protein of interest; it indicates which domains are making contact and provides a confidence score for each interaction.

Innoprot in Derio, Spain, provides an interaction service using tag-based purification designed for high-throughput analysis. And Invitrogen's ProtoArray Protein–Protein Interaction Service uses microarrays containing more than 9,000 human proteins to identify proteins that interact with any protein of interest. L.B.