Maturing interactions

The maturation of large-scale protein-protein interaction methodologies calls for improved methods to assess performance and data quality.

In December 2006 we published a special issue dedicated to protein-protein interaction (PPI) methods. Now almost exactly two years later, we again present a collection of research papers on this subject. The differences between then and now are noteworthy.

Just as sequencing has matured from a method targeted at individual genes to the present-day whole-genome efforts, PPI studies are shifting to proteome-wide experiments. Reflecting this shift, the papers in this issue deal with large-scale interaction studies and have the ultimate goal of determining the complete interactomes of organisms. But what will this require? Tellingly, a central emerging theme is the importance of evaluating the quality of interactions in large-scale datasets.

As part of its maturation, sequencing technology required adjustments in its methods for quality control. But for sequencing, the desired quality was easy to define and relatively easy to obtain, whereas data quality for PPI experiments is harder to measure, and defining a finished interactome is more challenging.

The data analysis methods described in the reports on pages 83 and 91 in this issue are an important step forward in the empirical assessment of PPI data quality. The positive and negative human PPI reference sets they provide are an experimentally testable open resource for quantifying the performance of PPI screens, and we hope these will be built upon and used by others. Just as the report on page 47 provides comparable reference datasets for *Caenorhabditis elegans*, it will be important to define reference sets for other organisms. The size of these datasets is ambitious but still small relative to the scale of the task at hand; care must therefore be exercised in their use. The negative reference set, for example, is too small to directly measure the false positive rate and false discovery rate of PPI assays.

In addition to providing empirical confidence assessments of PPI data, these quality measures allow estimation of total interactome size. Unfortunately, size estimates provide only a partial picture of the total experimental effort required to map an interactome. Modeling of various experimental strategies, as done on page 55, is therefore important for determining the magnitude of the mapping task ahead and how best to minimize the effort required for its completion. The results of this modeling show that probability thresholding—where each interaction is assigned an evidence-based probability that is continually adjusted—allows a substantial reduction in experimental effort, provided that the interaction assays have high complementarity (low overlap).

Going forward it will be important to determine the complementarity of different assays so that experimental designs can be optimized to provide the largest increase in interactome coverage with the least effort. A lingering concern is that one could fall victim to the illusion that combining complementary assays will necessarily provide high coverage—although in fact the high false negative rates of the assays mean that, just like the individual assays, their combination will never be able to interrogate some protein pairs. Regions of the interactome may thus remain hidden. For this reason it remains important to develop new assays that can sample new interactome space, such as the promising microfluidic approach described on page 71.

Although generation and evaluation of new data is critical, it is also important that the enormous body of interaction data already present in the existing literature be used productively. The published literature constitutes an incredibly valuable resource—one that is exploited to generate the high quality reference sets in this issue—but the level of evidence for each interaction is highly variable, and currently the curation process does not produce a quality score associated with each interaction to reflect this heterogeneity. The reports on pages 39 and 75 suggest that there is much room for improvement in overall curation quality and that properly performed large-scale screens can be of higher average quality.

If used properly, even low-quality literature-curated data that include a high rate of false positives and false negatives can improve the efficiency with which an interactome mapping project is completed, but, by improving curation quality, the existing collections could be made far more valuable. Efforts such as that on page 39 to integrate literature-curated data from multiple databases, and allow re-curation and filtering to improve data quality, are a useful step in this direction, but centralized efforts to improve curation and permit scoring of reported interactions would be a boon to all.

There is no doubt that large-scale PPI screening is in ascendance. The time for concerted efforts to establish clear standards and methods to assess assay performance and data quality is now.