Chapter XIII
Protein Interactions for Functional Genomics

Pablo Minguez
Centro de Investigación Príncipe Felipe (CIPF), Spain

Joaquin Dopazo
Centro de Investigación Príncipe Felipe (CIPF), Spain

ABSTRACT

Here the authors review the state of the art in the use of protein-protein interactions (ppis) within the context of the interpretation of genomic experiments. They report the available resources and methodologies used to create a curated compilation of ppis introducing a novel approach to filter interactions. Special attention is paid in the complexity of the topology of the networks formed by proteins (nodes) and pairwise interactions (edges). These networks can be studied using graph theory and a brief introduction to the characterization of biological networks and definitions of the more used network parameters is also given. Also a report on the available resources to perform different modes of functional profiling using ppi data is provided along with a discussion on the approaches that have typically been applied into this context. They also introduce a novel methodology for the evaluation of networks and some examples of its application.

INTRODUCTION

The available data for protein-protein interactions (ppis) has increased enormously in the last few years with the emergence of high-throughput techniques that can report thousands of ppis in a short time span. The most used techniques in this field are: yeast two hybrid (y2h), tandem affinity purification (TAP) and high-throughput Mass Spectrometry techniques (MS). Reviews on these and related methodologies can be found in Drewes and Bouwmeester (2003), Cho *et al*. (2003), Falk *et al*. (2007) and Berggard *et al*. (2007).

The reliability of this data is not exempt of controversy. Studies comparing resulting data from several experiments demonstrate that the overlap between them is not as complete as desirable. This can be due to the fact that some methodologies do not reach the saturation point (Bader & Hogue, 2002) or because of the lack of accuracy and coverage on some of them (von Mering *et al*., 2002). A conventional large-scale experiment can cover only 3-9% of the total interactome, so limited overlap should be expected (Han *et al*., 2005). False positives are also

Copyright © 2009, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.

a problem: in y2h these represent up to 50% of the total data (Ito *et al*., 2001; Mrowka *et al*., 2001). Moreover, there is a bias in the functional categories of the ppis each technique detects, e.g. y2h fails in detecting proteins involved in translation (von Mering *et al*., 2002).

Beyond discussions about accuracy and coverage of this kind of experiments, the relevance of ppis in the cellular machinery has fostered an unprecedented interest in the exploration of the interactome of model organisms such as *Saccharomyces cerevisiae* (Uetz *et al*., 2000; Ito *et al*., 2001), *Drosophila melanogaster* (Gio et al., 2003; Formstecher *et al*., 2005), *Caenorhabditis elegans* (Li *et al*., 2004) or human (Stelzl *et al*., 2005, Rual *et al*., 2005), just to cite a few examples.

Actually, after years of intensive study, there is a high-quality, literature curated set of ppis free from false positives that probably represents the complete yeast interactome (Reguly *et al*., 2006). In the case of human, the scenario is still far away from this degree of detail. The estimated size of the human interactome is of 650,000 ppis (Stumpf *et al*., 2008). None of the public databases contain more than 10% of this number of ppis, and a compilation of all the known ppis would only cover about 10% of the interactions.

The interactome is an abstract scaffold that does not provide information about particular conditions, cell developmental stage or cell type in which a particular ppi occurs (if any). To infer a case-specific interactome it is necessary to integrate other types of data that provide information that allows inferring the active ppis at a particular condition. To achieve this, the transcriptome, defined as the set of transcripts that are expressed at a given moment in a particular cell type, can be used. An integrative study of the interactome filtered by the transcriptome will provide valuable information on the active ppis in a given cell state.

Actually, ppis play a central role at almost every level of cell activity: they are involved in the structure of organelles (structural proteins), transport machinery (nuclear pore importins), response to stimulus (signalling cascades), regulation of gene expression (transcription factors), protein modification (kinases) among many other processes. The proper use of this type of information is of crucial importance in order to understand cell behaviour.

However, the conventional methodologies used to understand the functional basis of the cell behaviour are almost restricted to functional profiling methods. Such methods exploit the differences observed in the comparison of transcriptomes among different experimental conditions to find over-representations of predefined functional modules of genes (see Dopazo 2006 for a recent review). Classically, standard annotations like Gene Ontology (GO) terms (Ashburner et al., 2000), KEGG pathways (Kanehisa et al., 2004) or Mesh terms, have been used to define such modules. Nevertheless, ppis have not extensively used for such puroses.

The combination of expression and interactions has been used to infer gene function (Ideker *et al*., 2001), to extract signatures to predict disease phenotypes (Camargo & Azuaje, 2007; Lee *et al*., 2007; Liu *et al*., 2007; Chuang *et al*., 2007) as well as to detect possible drug targets by inferring topological features of particular classes of genes (Wachi *et al*., 2005; Johsson and Bates, 2006).

It is widely accepted that there are very few processes that can be explained by the action of a single protein. On the contrary, the units of activity involved in cellular processes seem to be modules composed by several interacting molecules (Hartwell *et al*., 1999; Barabasi and Oltvai, 2004). Apart from classical definitions of these modules, such as proteins that share a GO term or proteins integrating the same biological pathway, ppi data is also used to define modules, as representative of units of action characterized by the interaction of their components.

INTERACTOME-RELATED DEFINITIONS

PPI Resources

In this new era of massive production of biological data, an important challenge is its storage in a standardised format with the proper annotation. This facilitate further queries, as simple as possible, to retrieve relevant information from the databases. Data from high-throughput technologies, such as DNA sequences or microarray experiments have developed structured formats to submit the data to the databases with annotations following an ontology-based vocabulary. Learning from those experiences, the Proteomic Standards Initiative (PSI) of the

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/protein-interactions-functional-genomics/5567

Related Content

BioTextRetriever: A Tool to Retrieve Relevant Papers
Célia Talma Gonçalves, Rui Camachoand Eugénio Oliveira (2011). *International Journal of Knowledge Discovery in Bioinformatics (pp. 21-36).*
www.irma-international.org/article/biotextretriever-tool-retrieve-relevant-papers/63615

Introduction to Data Classification
(2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations (pp. 10-12).*
www.irma-international.org/chapter/introduction-data-classification/53894

New Trends in Graph Mining: Structural and Node-Colored Network Motifs
Francesco Bruno, Luigi Palopoliand Simona E. Rombo (2012). *Computational Knowledge Discovery for Bioinformatics Research (pp. 259-278).*
www.irma-international.org/chapter/new-trends-graph-mining/66715

Pattern Differentiations and Formulations for Heterogeneous Genomic Data through Hybrid Approaches
Arpad Kelemenand Yulan Liang (2006). *Advanced Data Mining Technologies in Bioinformatics (pp. 136-154).*
www.irma-international.org/chapter/pattern-differentiations-formulations-heterogeneous-genomic/4250

The Role of Genetic Data Analysis for Precision Therapy in Cancer: Personalized Medicine Concept in Cancer Treatment
Sunishtha Singh Yadav, Vandana Chauhan, Vijeta Singhand Svenja Kohler (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology (pp. 1467-1483).*
www.irma-international.org/chapter/role-genetic-data-analysis-precision/342585