BIOINFORMATICS APPLICATIONS NOTE Vol. 23 no. 18 2007, pages 2495–2497
doi:10.1093/bioinformatics/btm373

Systems biology

APID2NET: unified interactome graphic analyzer

Juan Hernandez-Toro, Carlos Prieto and Javier De Las Rivas*

Bioinformatics and Functional Genomics Research Group, Cancer Research Center (IBMCC-CIC, CSIC-USAL), Salamanca, Spain

Received on April 13, 2007; revised on June 7, 2007; accepted on July 11, 2007

Advance Access publication July 21, 2007

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Exploration and analysis of interactome networks at systems level requires unification of the biomolecular elements and annotations that come from many different high-throughput or small-scale proteomic experiments. Only such integration can provide a non-redundant and consistent identification of proteins and interactions. APID2NET is a new tool that works with Cytoscape to allow surfing unified interactome data by querying APID server (http://bioinfow.dep.usal.es/apid/) to provide interactive analysis of protein–protein interaction (PPI) networks. The program is designed to visualize, explore and analyze the proteins and interactions retrieved, including the annotations and attributes associated to them, such as: GO terms, InterPro domains, experimental methods that validate each interaction, PubMed IDs, UniProt IDs, etc. The tool provides interactive graphical representation of the networks with all Cytoscape capabilities, plus new automatic tools to find concurrent functional and structural attributes along all protein pairs in a network.

Availability: http://bioinfow.dep.usal.es/apid/apid2net.html

Contact: jrivas@usal.es

Supplementary information: Installation Guide and User's Guide are supplied at the Web site indicated above.

1 INTRODUCTION

The advancement of genome and proteome-wide experimental technologies have introduced modern biology in the high complexity of living cells, where thousands of biomolecules work together in an interactive way, with many short and long range cross-talks and cross-regulations. To achieve a first level of understanding of such cellular complexity we need to unravel the physical interactions that occur between all the proteins that integrate an active working cell. The compendium of all known protein–protein interactions (PPIs) for a given cell or organism is called the *interactome* and in recent years several key publications have used high-throughput proteomic techniques to determine the interactome in model organisms: yeast *Saccharomyces cerevisiae* (Ito *et al*., 2001; Uetz *et al*., 2000), fly *Drosophila melanogaster* (Formstecher *et al*., 2005; Giot *et al*., 2003) and worm *C.elegans* (Li *et al*., 2004). More recently several efforts have focused on the human interactome (Rual *et al*., 2005; Stelzl *et al*., 2005). The PPI data coming from many publications done with high-throughput proteomic techniques or with small-scale experimental methods are being collected and stored in several databases, like DIP (Xenarios *et al*., 2002) and IntAct (Hermjakob *et al*., 2004). Recently we have developed a bioinformatic web server called APID (Prieto and De Las Rivas, 2006), that integrates the interactome data of five main PPI databases unifying them in a common platform. At present, June-2007, APID includes 42 699 proteins from 15 organisms and a total of 156 688 interactions. It also includes information to help in the validation of the interactions, like the number of experimental methods that prove each interaction or the existence of structural domains that may interact according to iPfam. iPfam is a resource that describes domain–domain interactions that are observed in PDB entries (Finn *et al*., 2005).

The complete interactomes are rather complex systems so, despite that APID and other web tools include graphic displays to show sections of the interactome network, such graphical visualizations are quite limited when a specific study of a proteome subset or a protein family wants to be undertaken in detail. Moreover, many biomolecular research groups may demand specific PPIs selection to explore and analyze interactively such concrete network within the interactome landscape.

2 MINING THE INTERACTOME: CYTOSCAPE AND APID

Cytoscape is a bioinformatic software for visualizing molecular interaction networks and integrating these interactions with other biological data (http://www.cytoscape.org/). Cytoscape allows and promotes the integration of additional plugins that can provide network and profiling analyses, new layouts, connection with databases, etc. In recent years there have been several publications reporting stand-alone applications to investigate protein interaction networks, however most of them provide independent software that cannot be combined and whose development and maintenance sometimes may face uncertain future. The development of tools integrated in Cytoscape can provide useful and robust applications, since a large international community is collaborating writing advance software for this common and well-maintained platform.

Following these ideas we have developed a Java application called APID2NET that allows exploration, annotation and analysis of one or more subsets of the protein interactome. The tool is completely integrated in Cytoscape as a plugin and allows to query APID using a transparent Servlet interface and

*To whom correspondence should be addressed.

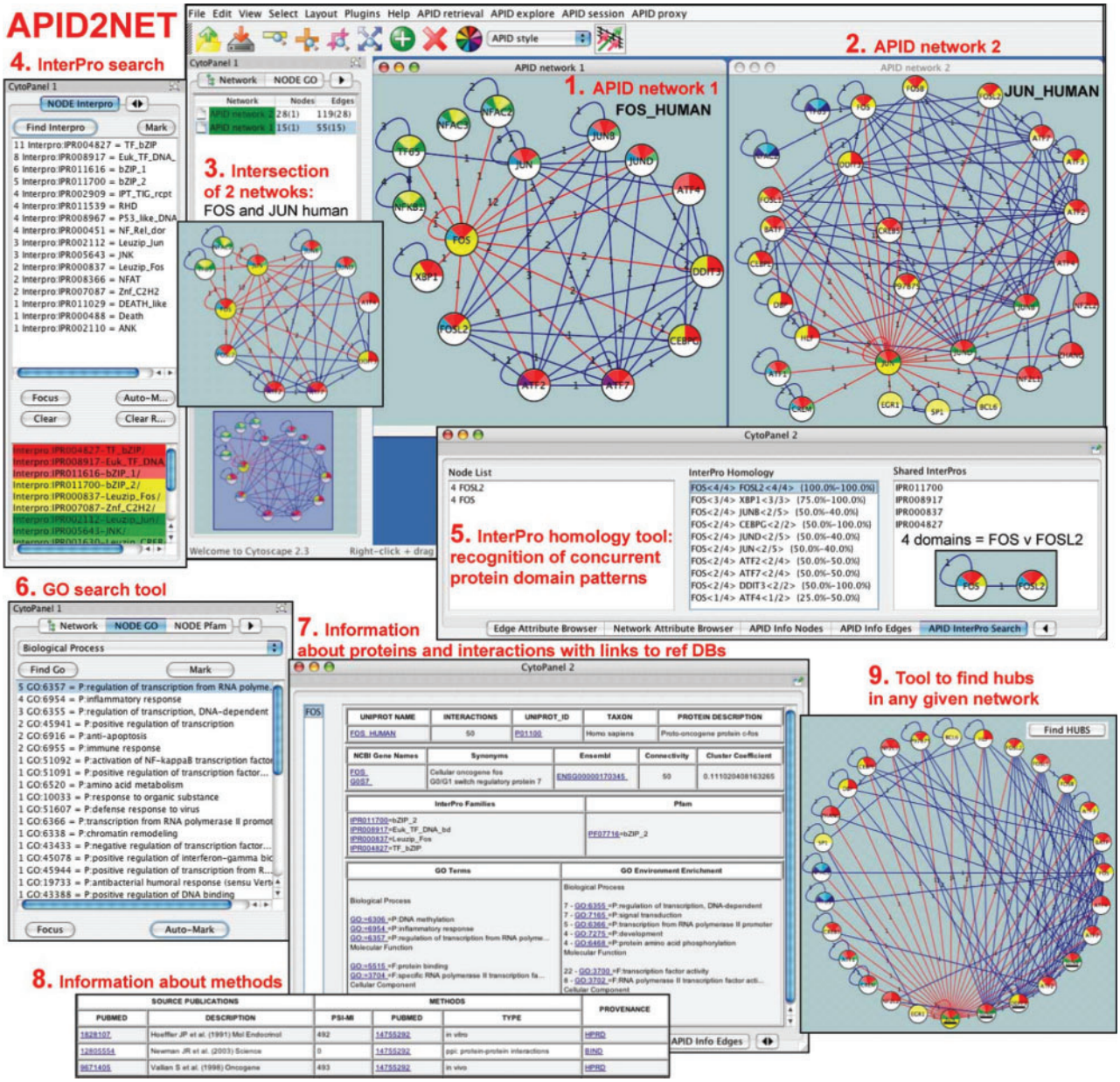

Fig. 1. Workflow of the application in nine consecutive windows. It includes an example of the protein interaction networks for two human proteins involved in transcription: FOS and JUN.

fetch the interaction network for a given protein or a given list of proteins. Once a network is downloaded it retains all the information present in APID, and includes several new features to explore and find co-occurrence of functional and structural annotations among the proteins.

3 FEATURES TO EXPLORE AND ANALYZE THE NETWORKS

APID2NET provides an interface within Cytoscape to allow on-line PPI data retrieval and further interactive analysis of nodes and edges. The main features included in the application are:

- 'APID retrieval' menu (to retrieve and visualize PPI networks):

(1) Search engine to retrieve PPI networks from APID server (i.e. query and fetch data). The search can retrieve the PPI network of a single protein or a protein list.

(2) The PPI networks can be retrieved using quality filters, like: a minimal number of experimental methods that

validate the interactions; the presence of iPfam predicted interacting domains.

(3) Simple clicking tool to expand the network from any selected node and retrieve new interactions.

- 'APID explore' menu (to explore functional and structural annotations and to locate hubs):

(1) Each network can be explored using several protein attributes: GO terms (within the three categories BP, MF or CC), Pfam sequence-based domains, InterPro domains and motifs. Each search displays the attributes ordered by the number of times that appear in the query interaction network ('CytoPanel-1').

(2) The information about each protein (node) and each interaction (edge) is displayed in tables ('CytoPanel-2'), including the experimental methods that validate each interaction with the PubMed literature references. The tables include links-out to the source files in public databases (UniProt, PubMed, GO, etc). The nodes and edges in the tables are interactive to allow easy location in the network window.

(3) 'Find' and 'Auto-Mark' (in 'CytoPanel-1' submenus): to find annotations and attributes about nodes or edges. Functional and structural attributes (GO, Pfam, InterPro) that are concurrent in at least one protein-pair can be found and color-coded using 'Auto-Mark'.

(4) 'Hubs' tool ('CytoPanel-1', 'Mark-NODE'): to locate the hubs present in a given network. The tool uses an algorithm that searches for the protein nodes that include a given percentage of the non-repetitive edges in the network (65% by default).

(5) InterPro homology tool ('CytoPanel-2', 'APID InterPro Search'): to find pair-wise protein domain pattern homology based on InterPro. The tool searches for all possible protein-pairs in a network and marks the domains and motifs from InterPro that are common to each pair. The number of common domains is counted providing a ranking from more to less similar pairs.

- 'APID session' menu (to save and reload data):

(1) Designed to save on the local PC a complete working session with the information and analyses that had been done by a user.

All the tools and features are explained in detail in the User's Guide available on-line.

4 EXAMPLE: EXPLORE AND COMPARE FOS AND JUN HUMAN NETWORKS

An example showing the main features of APID2NET is included in Figure 1, that presents the interaction networks of two human proteins involved in transcription: FOS_HUMAN (P01100) and JUN_HUMAN (P05412). The networks are filtered to retrieve only the proteins that have iPfam validation. In this way, FOS network includes 15 nodes and 28 edges; and JUN network includes 28 nodes and 119 edges. The FOS–JUN interaction is well known and it is validated by 12 reported methods. The protein annotation co-occurrence discovery tool that uses InterPro, provides a clear way to find protein homologous. Window 5, in Figure 1, shows the interaction FOS–FOSL2: FOSL2 is a FOS-related antigen and presents the same domain architecture as FOS. The InterPro co-occurrence analysis also shows that proteins JUN, JUNB and JUND have the same domains pattern, and in this way they are found putative paralogous proteins. The described co-occurrence tool can help to unravel functional protein relationships and to improve our understanding of the interactome modules and architecture.

ACKNOWLEDGEMENTS

Funding to pay the Open Access publication charges was provided by the Spanish Ministery of Health (MSyC, ISCIII, FIS grant reference PI061153) and Junta Castilla y Leon (grant reference CSI03A06).

Conflict of Interest: none declared.

REFERENCES

Finn,R.D. *et al.* (2005) iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.

Formstecher,E. *et al.* (2005) Protein interaction mapping: a Drosophila case study. *Genome Res.*, **15**, 376–384.

Giot,L. *et al.* (2003) A protein interaction map of Drosophila melanogaster. *Science*, **302**, 1727–1736.

Hermjakob,H. *et al.* (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res.*, **32**, D452–455.

Ito,T. *et al.* (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA*, **98**, 4569–4574.

Li,S. *et al.* (2004) A map of the interactome network of the metazoan C. elegans. *Science*, **303**, 540–543.

Prieto,C. and De Las Rivas,J. (2006) APID: Agile Protein Interaction DataAnalyzer. *Nucleic Acids Res.*, **34**, W298–W302.

Rual,J.F. *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173–1178.

Stelzl,U. *et al.* (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957–968.

Uetz,P. *et al.* (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.

Xenarios,I. *et al.* (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.*, **30**, 303–305.