BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 14 2012, pages 1949–1950
doi:10.1093/bioinformatics/bts289

Databases and ontologies

Advance Access publication May 15, 2012

IRView: a database and viewer for protein interacting regions

Shigeo Fujimori[1], Naoya Hirai[1], Kazuyo Masuoka[1], Tomohiro Oshikubo[1,2], Tatsuhiro Yamashita[1,3], Takanori Washio[1,4], Ayumu Saito[5], Masao Nagasaki[5], Satoru Miyano[5] and Etsuko Miyamoto-Sato[1,5,*]

[1]Division of Interactome Medical Sciences, Institute of Medical Science, The University of Tokyo, Tokyo 108-8039, Japan, [2]Production Solution Business Unit, Production Solution Division.II, Solution Department I, Fujitsu Advanced Engineering Ltd., Tokyo 163-1017, Japan, [3]BioIT Business Development Unit, Fujitsu Ltd., Chiba 261-8588, Japan, [4]Bioinformatics Department, RIKEN GENESIS Co., Ltd. Yokohama 230-0045, Japan and [5]Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8039, Japan

Associate Editor: Jonathan Wren

ABSTRACT

Summary: Protein–protein interactions (PPIs) are mediated through specific regions on proteins. Some proteins have two or more protein interacting regions (IRs) and some IRs are competitively used for interactions with different proteins. IRView currently contains data for 3417 IRs in human and mouse proteins. The data were obtained from different sources and combined with annotated region data from InterPro. Information on non-synonymous single nucleotide polymorphism sites and variable regions owing to alternative mRNA splicing is also included. The IRView web interface displays all IR data, including user-uploaded data, on reference sequences so that the positional relationship between IRs can be easily understood. IRView should be useful for analyzing underlying relationships between the proteins behind the PPI networks.

Availability: IRView is publicly available on the web at http://ir.hgc.jp/.

Contact: nekoneko@ims.u-tokyo.ac.jp

Received on November 21, 2011; revised on March 24, 2012; accepted on May 10, 2012

1 INTRODUCTION

Protein–protein interactions (PPIs) and their networks play central roles in governing cellular processes. Recently, much effort has been put in to collecting binary interaction data (e.g. Rual *et al*., 2005) to dissect PPI networks. These interaction data have been compiled in public biomolecular interaction databases. For example, BioGRID (Breitkreutz *et al*., 2008) and IntAct (Aranda *et al*., 2009) are major molecular interaction databases of PPIs. These databases primarily contain PPI data at the protein level, namely pairs of protein names. Two recent articles, one of them from our group, have reported large-scale experimental data on region- or domain-based protein interactions determined by high-throughput methods in human (Miyamoto-Sato *et al*., 2010) and in *Caenorhabditis elegans* (Boxem *et al*., 2008). Usually, proteins interact with other proteins through regions (the interacting regions, IRs) that are specific to each interaction; therefore, simultaneous interactions with multiple proteins are possible (Kim *et al*., 2006). Furthermore, some IRs are competitively involved in interactions with different proteins. To comprehend the complicated relations underlying PPIs in more detail, further refined interaction data are required.

In this article, we describe IRView, a database and viewer for IRs of proteins, which focuses on the regions required for PPIs. IRView contains, as a primary data source, IRs that were determined using the *in vitro* virus (IVV) method (human data: Miyamoto-Sato *et al*., 2010; mouse data: Horisawa *et al*., 2004 and Miyamoto-Sato *et al*., 2005) and the yeast two-hybrid (Y2H) method. DOMINO (Ceol *et al*., 2007) is a database of domain–domain interactions that is similar in scope to IRView. DOMINO stores data on IRs described in the scientific literature and applies the existing domain/motif names from the InterPro (Hunter *et al*., 2009) database to each of the IRs. The IR data in IRView include InterPro domain/motif regions but are not restricted to the InterPro annotations. Users can also compare IRs with variable sites susceptible to non-synonymous single nucleotide polymorphisms (nsSNPs) and variable regions arising from alternative mRNA splicing. IRView also supports a viewer that allows users to compare the positional relationships of IRs in protein reference sequences and in 3D structures (when available). IRView should be useful for investigating the hidden relationships between the proteins behind protein interaction networks.

2 CONTENTS AND FEATURES

2.1 The IR and other functional region data

The current version of IRView contains 3417 unique IRs as the default IR data. The IR data correspond to 1901 genes and were obtained using the IVV (human data: Miyamoto-Sato *et al*., 2010; mouse data: Horisawa *et al*., 2004 and Miyamoto-Sato *et al*., 2005) and Y2H (Sugaya *et al*., 2007) methods. Over half the IR data (2629 IRs) were derived from the results obtained using the IVV method (Miyamoto-Sato *et al*., 2010). Although the current data and sources for IRView are limited, we plan to add our original, experimental data and data extracted from the literature. The IR data can be downloaded in the PSI-MITAB format (Kerrien *et al*., 2007) file.

The conserved domains and motifs data that were used to visualize the positional relationships of the IRs were retrieved from the InterPro database (Hunter *et al*., 2009). Data on the nsSNPs that can lead to amino acid changes and potentially affect protein interactions (Mendelsohn, 2004; Schuster-Bockler and Bateman, 2008) were

*To whom correspondence should be addressed.

obtained from the dbSNP (Smigielski *et al*., 2000). Variable regions derived from alternative mRNA splicing that may potentially affect protein interactions (Resch *et al*., 2004) were defined based on the results of pair-wise alignments between the various isoforms. The 3D structure data were downloaded from the Protein Data Bank. When 3D models of complexes were available, information about the interacting amino acid residues on different peptide chains (defined as amino acids that were within a distance <4.0 Å of each other) were also added to the IR data.

2.2 Reference sequence-centered map

One of the main features of IRView is that all positional data are standardized to positions in reference sequences. IRView uses the NCBI RefSeq sequences as the reference protein sequences. When different isoforms of a protein are recorded in the RefSeq database, the longest sequence was selected as the representative sequence (RS) and the others were treated as related sequences. Standardization of the position data was achieved by pair-wise alignments between the RS and the other related sequences using ClustalW 2.0 (Larkin *et al*., 2007). As a result of this standardization, users can easily capture the positional relationships between independently annotated regions from different sources. Using the standardized position data, IRView can provide information about the positional relationships between the IRs in one protein sequence that interact with different proteins. In addition, IRView provides information on the positional relationships between IRs and other annotated regions (e.g. InterPro regions). Whether or not an IR overlaps with any other annotated region is indicated by special icons that accompany each IR entry. Of the 3417 IRs in the current version of IRView, 1492 IRs overlap with known domain/motif regions, 521 IRs overlap with nsSNPs, 207 IRs overlap with structured regions, 102 IRs overlap with variable regions derived from alternative mRNA splicing and 691 IRs overlap with other IRs.

3 DESCRIPTION OF THE IRVIEW INTERFACE

IRView supports searches by protein name, gene symbol, NCBI Entrez GeneID, RefSeq ID, species name and free keywords. IRView also supports the use of field specifiers to limit the scope of searches. Query results are returned as a list of RSs which correspond to a 'Region information' page consisting of a number of sections: a 'Gene summary' section for basal information on the RSs; a 'Protein sequences and regions' section which contains positional information for the IRs and other annotated regions related to the RS; and a 'Custom regions' section which allows users to compare positional relationships between arbitral IRs (e.g. in-house data) and the IRs in the database. The 'Protein sequences and regions' section is divided into several subsections: 'Representative sequence', 'Related sequences', 'Structured regions', 'Domain/Motif regions', 'Variable regions', 'Non-synonymous SNPs', 'Interacting regions' and 'Contacting amino acids'. To make it easier to compare two or more annotated regions (e.g. comparing variant regions with IRs to infer the impact of alternative splicing), unnecessary subsections can be collapsed. Details of these subsections are described in the Help page of IRView.

IRView also possesses a system for mapping specific regions to 3D structure(s) when the corresponding structure data are available. Users can map regions of interest to the 3D structure individually or simultaneously via the in-lined Jmol applet (http://www.jmol.org/) on any Java-enabled web browser.

ACKNOWLEDGEMENTS

We thank Katsuya Hino for his support in constructing the database. We also thank Dr Hiroshi Yanagawa for helpful discussions and advice.

Funding: Grant-in-Aid for Scientific Research on Innovative Areas 'Integrative Systems Understanding of Cancer for Advanced Diagnosis, Therapy and Prevention (No. 4201)' (grant number 23134510) of The Ministry of Education, Culture, Sports, Science and Technology, Japan (to S.F.); Female Researcher Science Grant from Shiseido Co., Ltd (to E.M.-S.).

Conflict of Interest: none declared.

REFERENCES

Aranda,B. *et al.* (2009) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.

Boxem,M. *et al.* (2008) A protein domain-based interactome network for C. elegans early embryogenesis. *Cell*, **134**, 534–545.

Breitkreutz,B.J. *et al.* (2008) The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.

Ceol,A. *et al.* (2007) DOMINO: a database of domain-peptide interactions. *Nucleic Acids Res.*, **35**, D557–D560.

Horisawa,K. *et al.* (2004) In vitro selection of Jun-associated proteins using mRNA display. *Nucleic Acids Res.*, **32**, e169.

Hunter,S. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **37**, D211–D215.

Kerrien, S. *et al.* (2007) Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.

Kim,P.M. *et al.* (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. *Science*, **314**, 1938–1941.

Larkin,M.A. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Mendelsohn,A.R. (2004) Interaction trap/two-hybrid system to identify loss-of-interaction mutant proteins. *Curr. Protoc. Mol. Biol.*, **Chapter 20**, Unit 20 28.

Miyamoto-Sato,E. *et al.* (2005) Cell-free cotranslation and selection using in vitro virus for high-throughput analysis of protein-protein interactions and complexes. *Genome Res.*, **15**, 710–717.

Miyamoto-Sato,E. *et al.* (2010) A comprehensive resource of interacting protein regions for refining human transcription factor networks. *PLoS ONE*, **5**, e9289.

Resch,A. *et al.* (2004) Assessing the impact of alternative splicing on domain interactions in the human proteome. *J. Proteome. Res.*, **3**, 76–83.

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

Schuster-Bockler,B. and Bateman,A. (2008) Protein interactions in human genetic diseases. *Genome Biol.*, **9**, R9.

Smigielski,E.M. *et al.* (2000) dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.*, **28**, 352–355.

Sugaya,N. *et al.* (2007) An integrative in silico approach for discovering candidates for drug-targetable protein–protein interactions in interactome data. *BMC Pharmacol.*, **7**, 10.