IRView: a database and viewer for protein interacting regions

Shigeo Fujimori1, Naoya Hirai1, Kazuyu Masuko1, Tomohiro Oshikubo1,2, Tatsuhiro Yamashita1,3, Takanori Washio1,4, Ayumu Saito5, Masao Nagasaki5, Satoru Miyano5 and Etsuko Miyamoto-Sato1,5,*

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

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 (Breitkreute 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., 2008) 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.

1 To whom correspondence should be addressed.

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., 2004) 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...
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 mSNPs, 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 provides 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


S.Fujimori et al.