A Tutorial on Path2PPI Package: Predicting Protein-Protein Interactions for a Set of Proteins of Interest in Non-model Organisms.

Proteins are versatile biomolecules that are central to eukaryotic cell complexity. Understanding protein-protein interactions (PPIs) may lead to discovery of novel proteins that have not been characterized as well as a more thorough comprehension of PPIs in disease states that may aid for new drug discovery. While there now exists a variety of PPI databases such as INTACT, BIOGRID, and STRING, they harbor predicted PPIs for only model organisms (human, mouse, rat, yeast, E.coli, C.elegans and D.melanogaster). Most PPIs predictions are results from gene ontology functional annotation or genetic co-expression data and there are little to no databases for PPIs for non-model organisms where only sequenced data is available experimentally or in genomic databases. To predict a biological process of interest’s proteins and their interactions in non-model organisms, we will employ the usage of Path2PPI package in R studio. Utilization of this package requires 3 types of data: a set of protein from a corresponding biological process from a model organism to be used as reference proteins, a *iRefindex* data frame that contains relevant PPIs of the first dataset, and a NCBI BLAST+ homology dataset that contains blast data of target non-model organism’s proteomic data searched against the reference specie or species proteomic data. The function predictPPI in the package creates predicted PPIs by searching relevant PPIs from *iRefindex* PPI dataset for the reference proteins and checks the homologous NCBI BLAST homology datasets for the generated relevant PPI interactions. Should a homologous proteins to the target species protein makes the homology score, the algorithm will automatically pick it up as a predicted interaction. The package can be utilized further to plot the predicted PPI as path2ppi has igraph package integrated in it as well. For this tutorial, we will rest at the plotting of the predicted PPI interaction. But Path2PPI package can also further be used to plot the predicted PPIs in more detail such as to plot from which reference species the different predicted interactions originated as well to view which proteins have homologous relationship or direct interaction as well.

References

1. Philipp O, Osiewacz HD, Koch I. Path2PPI: an R package to predict protein-protein interaction networks for a set of proteins. Bioinformatics. 2016 May 1;32(9):1427-9. doi: 10.1093/bioinformatics/btv765. Epub 2016 Jan 5. PMID: 26733452; PMCID: PMC4848400.
2. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421. doi: 10.1186/1471-2105-10-421. PMID: 20003500; PMCID: PMC2803857.
3. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007;2(10):2366-82. doi: 10.1038/nprot.2007.324. PMID: 17947979; PMCID: PMC3685583.
4. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006;1695(5), 1-9.
5. Razick S, Magklaras G, Donaldson IM. iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics. 2008 Sep 30;9:405. doi: 10.1186/1471-2105-9-405. PMID: 18823568; PMCID: PMC2573892.
6. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014 Jan;42(Database issue):D199-205. doi: 10.1093/nar/gkt1076. Epub 2013 Nov 7. PMID: 24214961; PMCID: PMC3965122.