Systems Biology

IMMAN: an R/Bioconductor package for Interolog protein network reconstruction, Mapping and Mining ANalysis

Payman Nickchi[1,2,7], Soheil Jahangiri-Tazehkand[3,4,7], Minoo Ashtinai[5,7], Abdollah Safari[2*], Mehdi Mirzaie[6*] & Mohieddin Jafari[1,4*]

[1]Drug Design and Bioinformatics Unit, Medical Biotechnology Department, Biotechnology Research Center, Pasteur Institute of Iran, 69, Pasteur St, 13164, Tehran, Iran, [2]Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, V5A 1S6, Burnaby, BC, Canada, [3]Department of Computer Science, Shahid Beheshti University, Evin, Tehran, Iran. [4]School of Biological Science, Institute for Research in Fundamental Sciences (IPM), Shahid Lavasani St, PO Box 19395-5746, Tehran, Iran, [5]Department of Computer Science and Statistics, Faculty of Mathematics, K. N. Toosi University of Technology, P.O. Box 16315-1618, Tehran, Iran, [6]Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Jalal Ale Ahmad Highway, PO Box 14115-134, Tehran, Iran, [7]These authors have contributed equally to this work.

*To whom correspondence should be addressed.

Summary

IMMAN is a software for reconstructing Interolog Protein Network (IPN) by integrating several Protein-protein Interaction Networks (PPIN). Users can unify different PPINs to mine conserved common network among species. IMMAN helps to retrieve IPNs with different degrees of conservation to engage for protein function prediction analysis based on protein networks.

Availability: IMMAN is freely available at https://bioconductor.org/packages/IMMAN, http://jafarilab.com/imman/ .

Contact: mirzie@ipm.ir, mjafari@ipm.ir, asafari@sfu.ca
Supplementary information: Supplementary data are available online.

1 Introduction

Nowadays, technologies have provided access to tremendous amount of interactions at the molecular level. The study of these interactions, interactome, endeavor to model cellular and molecular processes (1, 2). Among these interactions, protein-protein interactions (PPI) are remarkable due to providing functional and structural description of executive molecules i.e. proteins. Nevertheless, PPI detection and prediction technologies are still entangling with reducing false-positive and -negative interactions (3-5). Accordingly, data integration is the best solution overall in spite of the improvement of experimental and computational methods,. STRING (6), BioNetBuilder Cytoscape app (7), IMP 2.0 (8), PINALOG (9), HIPPIE (10) and BIPS (11) are using this solution to reconstruct and refine PPI networks (PPIN). In other works, an evolutionarily conserved network with communal nodes and less false-positive links, Interolog Protein Network (IPN), was introduced as a benchmark for the evaluation of clustering algorithms (12). IPN clears up the arisen and remained interactions during evolution and help to excavate the remnants of ancestor PPIN (12-16). In this study, we present IMMAN, a software to integrate several PPINs and mine IPNs. IMMAN is free and is available as a Java program and an R/Bioconductor package.

2 Methods

IMMAN enables users to define two to four arbitrarily lists of proteins (by UniProt accession number) as inputs, and seek for evolutionarily conserved interactions in the integrated PPIN or IPN as an output. Briefly speaking, the method takes the following steps to accomplish this goal.

Step 1. First, the amino acid sequence of each input protein list is automatically retrieved from UniProt database.

Step 2. In the second step, IMMAN infers the orthologous proteins. To this end, the Needleman-Wunsch algorithms is employed to compute the pairwise sequence similarities. The reciprocal best hits are retrieved and applied in the next step to increase chance of orthologous pairs discovery. The user can adjust different parameters of alignment algorithm as well as the sequence similarity cutoff for orthology detection.

Step 3. In this step, the nodes of the IPN are specified. Each node of the network is defined as a set of mutually orthologous proteins (OPS) such that eachOPS belongs to a set of species involved in the analysis.

Step 4. In the fourth step, for each species, we extract singly the PPINs according to the proteins constitute the OPSs or IPN nodes. The PPINs are retrieved from STRING database. The user can adjust the minimal confidence score of STRING networks.

Step 5. Finally, the edges of the interolog network are computed. To this end, for every OPS pair, we count the number protein pairs (p_{ik}, p_{jk}) such that $p_i$and $p_j$are connected in the PPIN of species k. If this number exceeds a predefined cutoff (coverage cutoff), there would be an edge between the aforementioned nodes. The coverage cutoff can be also adjusted by the user to tune conservedness.

3 Results

After running IMMAN, the node list and the edge list of inferred IPN is produced. Additionally, IMMAN outputs the graphical representation of the network. The graphical output of IMMAN are produced using GraphViz (17) and igraph (18) in Java and R applications, respectively. The graphical representation of IMMAN on a sample dataset is depicted in Fig. 1. The sample dataset is available in Supplementary Data.

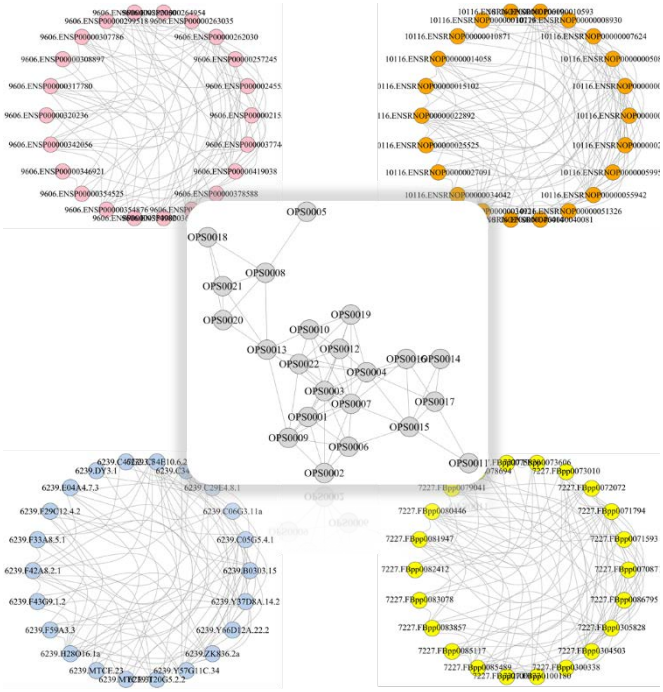

Fig1. The IPN of a sample datasets derived from four species namely; *H. sapiens* (top-left), *M. musculus* (top-right), *D. melanogaster* (bottom-left) and *C.elegans* (bottom-rigth).

Although, the size of IPN is tunable by several thresholds, but obviously, missing the edges in IPN is the cost of true positive discovery which is an ideal within PPI studies with inherent inconsistency (5, 19). However, function prediction is a prominent question in molecular biology and this approach pave its way based on evolutionary mechanism (20).

Acknowledgements

The authors would like to thank Dr. Mehdi Sadeghi for his valuable comments and discussions.

Funding

This work has been supported by the grant number No. BS 1395_0_01 provided by the school of biological sciences, Institute for Research in Fundamental Sciences, Tehran, Iran.

Conflict of Interest: The authors have no conflict of interest.

References

1. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, et al. A Proteome-Scale Map of the Human Interactome Network. Cell. 2014;159(5):1212-26.

2. Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease. Cell. 2011;144(6):986-98.

3. Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, et al. Literature-curated protein interaction datasets. Nature Methods. 2009;6(1):39-46.

4. Hart GT, Ramani AK, Marcotte EM. How complete are current yeast and human protein-interaction networks ? Genome Biology. 2006.

5. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, Yu H, et al. An experimentally derived confidence score for binary protein-protein interactions. Nat Methods. 2009;6(1):91-7.

6. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Research. 2016:gkw937-gkw.

7. Avila-Campillo I, Drew K, Lin J, Reiss DJ, Bonneau R. BioNetBuilder: automatic integration of biological networks. Bioinformatics (Oxford, England). 2007;23(3):392-3.

8. Wong AK, Park CY, Greene CS, Bongo La, Guan Y, Troyanskaya OG. IMP: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. Nucleic acids research. 2012;40(Web Server issue):W484-90.

9. Phan HTT, Sternberg MJE. PINALOG: a novel approach to align protein interaction networks--implications for complex detection and function prediction. Bioinformatics (Oxford, England). 2012;28(9):1239-45.

10. Schaefer MH, Fontaine JF, Vinayagam A, Porras P, Wanker EE, MA A-N. HIPPIE: Integrating protein interaction networks with experiment based quality scores. Plos One. 2012;7(2).

11. Garcia-Garcia J, Schleker S, Klein-Seetharaman J, Oliva B. BIPS: BIANA Interolog Prediction Server. A tool for protein-protein interaction inference. Nucleic Acids Res. 2012;40(Web Server issue):W147-51.

12. Jafari M, Mirzaie M, Sadeghi M. Interlog protein network: an evolutionary benchmark of protein interaction networks for the evaluation of clustering algorithms. BMC bioinformatics. 2015;16(1):319-.

13. Jafari M, Sadeghi M, Mirzaie M, Marashi S-A, Rezaei-Tavirani M. Evolutionarily conserved motifs and modules in mitochondrial protein–protein interaction networks. Mitochondrion. 2013;13:7.

14. Matthews LR. Identification of Potential Interaction Networks Using Sequence-Based Searches for Conserved Protein-Protein Interactions or "Interologs". Genome Research. 2001;11(12):2120-6.

15. Nguyen PV, Srihari S, Leong HW. Identifying conserved protein complexes between species by constructing interolog networks. BMC Bioinformatics. 2013;14(Suppl 16):S8-S.

16. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, et al. Protein interaction mapping in C. elegans using proteins involved in vulval development. Science (New York, NY). 2000;287(5450):116-22.

17. Ellson J, Gansner E, Koutsofios L, North SC, Woodhull G, editors. Graphviz—open source graph drawing tools. International Symposium on Graph Drawing; 2001: Springer.

18. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006;1695(5):1-9.

19. Shin CJ, Davis MJ, Ragan MA. Towards the mammalian interactome: Inference of a core mammalian interaction set in mouse. Proteomics. 2009;9(23):5256-66.

20. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome Res. 2004;14(6):1107-18.