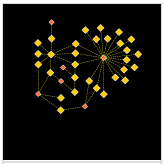

Jun 10, 2021

Protein interaction network analysis for Mendelian diseases

Nikoleta Vavouraki[1,2], James E. Tomkins[1], Eleanna Kara[3], Henry Houlden[4], John Hardy[3,5,6,7,8,9], Marcus J. Tindall[2,10], Patrick A. Lewis[11,3,1,9], Claudia Manzoni[12,1]

[1]Department of Pharmacy, University of Reading, Reading, RG6 6AX, United Kingdom;
[2]Department of Mathematics and Statistics, University of Reading, Reading, RG6 6AX, United Kingdom;
[3]Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, London, WC1N 3BG, United Kingdom;
[4]Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, WC1N 3BG, United Kingdom;
[5]UK Dementia Research Institute at UCL and Department of Neurodegenerative Disease, UCL IoN, UCL London, W1T 7NF United Kingdom;
[6]Reta Lila Weston Institute, UCL IoN, 1 Wakefield Street, London, WC1N 1PJ, United Kingdom;
[7]UCL Movement Disorders Centre, Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, WC1N 3BG, United Kingdom;
[8]Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China;
[9]Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD, 20815;
[10]Institute of Cardiovascular and Metabolic Research, University of Reading, Reading, RG6 6AS, United Kingdom;
[11]Department of Comparative Biomedical Sciences, Royal Veterinary College, London, NW1 0TU, United Kingdom;
[12]School of Pharmacy, University College London, London, WC1N 1AX, United Kingdom Contact: c.manzoni@ucl.ac.uk

 dx.doi.org/10.17504/protocols.io.buwanxae

c.manzoni

ABSTRACT

This protocol describes the steps to use experimentally validated human data to create a protein-protein interaction network (PPIN) based on disease causative genes. Network analysis (combination of topological functional analyses) will lead to the identification of biological processes relevant for disease and disease endophenotypes.

DOI

dx.doi.org/10.17504/protocols.io.buwanxae

EXTERNAL LINK

http://www.reading.ac.uk/bioinf/PINOT/PINOT_form.html

PROTOCOL CITATION

Nikoleta Vavouraki, James E. Tomkins, Eleanna Kara, Henry Houlden, John Hardy, Marcus J. Tindall, Patrick A. Lewis, Claudia Manzoni 2021. Protein interaction network analysis for Mendelian diseases. **protocols.io** https://dx.doi.org/10.17504/protocols.io.buwanxae

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Integrating protein networks and machine learning for disease stratification in the Hereditary Spastic Paraplegias. Vavouraki, N; Tomkins, JE; Kara, E; Houlden, H; Hardy, J; Tindall, MJ; Lewis, PA iScience 2021 Volume 24, Issue 5, 21 May 2021, 102484 Available online 28 April 2021. Published: May 21, 2021 doi: 10.1016/j.isci.2021.102484
Stratification of candidate genes for Parkinson's disease using weighted protein-protein interaction network analysis. Ferrari R, Kia DA, Tomkins JE, Hardy J, Wood NW, Lovering RC, Lewis PA, Manzoni C.BMC Genomics. 2018 Jun 13;19(1):452. doi: 10.1186/s12864-018-4804-9. PMID: 29898659
Weighted Protein Interaction Network Analysis of Frontotemporal Dementia. Ferrari R, Lovering RC, Hardy J, Lewis PA, Manzoni C.J Proteome Res. 2017 Feb 3;16(2):999-1013. doi: 10.1021/acs.jproteome.6b00934. Epub 2017 Jan 12. PMID: 28004582

Citation: Nikoleta Vavouraki, James E. Tomkins, Eleanna Kara, Henry Houlden, John Hardy, Marcus J. Tindall, Patrick A. Lewis, Claudia Manzoni (06/10/2021). Protein interaction network analysis for Mendelian diseases. https://dx.doi.org/10.17504/protocols.io.buwanxae

KEYWORDS

Protein-protein interaction, Functional stratification, Phenotypes, Causative genes, Topological analysis, Functional annotation, Hereditary Spastic Paraplegias.

CREATED

May 11, 2021

LAST MODIFIED

Jun 10, 2021

OWNERSHIP HISTORY

May 11, 2021 Urmilas

Jun 01, 2021 Oke Avwenagha

Jun 10, 2021 c.manzoni

PROTOCOL INTEGER ID

49826

Generation of PPI network

1 Select the protein products of genes relevant for disease (seeds).

2

Use the selected seeds as input to query the PINOT webtool (Tomkins et al., 2020) [http://www.reading.ac.uk/bioinf/PINOT/PINOT_form.html] or any other webtool that retrieves protein interactions from literature.

3 Screen the PPIs (protein-protein interactions) provided from PINOT to remove PPIs with a final score <3.

4 Score each node in the network based on the number of seeds to which it connects.

5 Select all nodes interacting with more than one seed.

6 Extract a subnetwork composed of the nodes selected in step 5. and all the connected seeds.

Citation: Nikoleta Vavouraki, James E. Tomkins, Eleanna Kara, Henry Houlden, John Hardy, Marcus J. Tindall, Patrick A. Lewis, Claudia Manzoni (06/10/2021). Protein interaction network analysis for Mendelian diseases. https://dx.doi.org/10.17504/protocols.io.buwanxae

Enrichment analyses

7 

Evaluate the functions associated with the PPIN using Gene Ontology terminology (Biological Processes [BPs] and Cellular Components [CCs] Gene Ontology [GO] annotations).

8 Evaluate the consistency of the results using 3 independent online tools, g:Profiler (RRID:SCR_006809) [https://biit.cs.ut.ee/gprofiler/gost], Gene Ontology using Panther's tool (RRID:SCR_004869) [http://pantherdb.org/] and WebGestalt (RRID:SCR_006786) [http://www.webgestalt.org/].

9 Group the enriched BP and CC GO terms by semantic similarity into semantic classes.

10 Remove the GO terms classified in the semantic classes "general" and "metabolism" as they refer to GO terms that provide limited functional specificity to the analysis.

11 Cluster the semantic classes into functional blocks and location blocks, respectively.

PCA & Hierarchical clustering

12 

In order to compare functional enrichment profiles, retrieved for different parts of the network, reduce data dimension for example conducting Principal Component Analysis (PCA) through R (RRID:SCR_001905, v. 4.0.2) using the prcomp() function of the stats package.

13 Conduct the hierarchical clustering using the hclust() function (R stats package) using Euclidean as a distance measure for row clustering.

14 Choose the best fit for the number of clusters derived from Hierarchical clustering based on the Silhouette method (Rousseeuw, 1987) and the Multiscale bootstrap resampling method (Suzuki and Shimodaira, 2006) (R package "pvclust").

Citation: Nikoleta Vavouraki, James E. Tomkins, Eleanna Kara, Henry Houlden, John Hardy, Marcus J. Tindall, Patrick A. Lewis, Claudia Manzoni (06/10/2021). Protein interaction network analysis for Mendelian diseases. https://dx.doi.org/10.17504/protocols.io.buwanxae