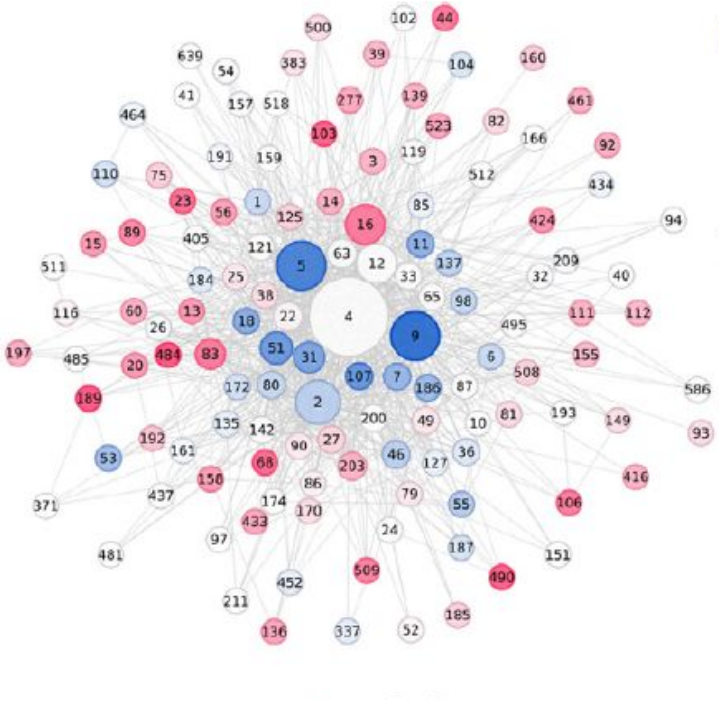
Functional profiling:

Network analysis



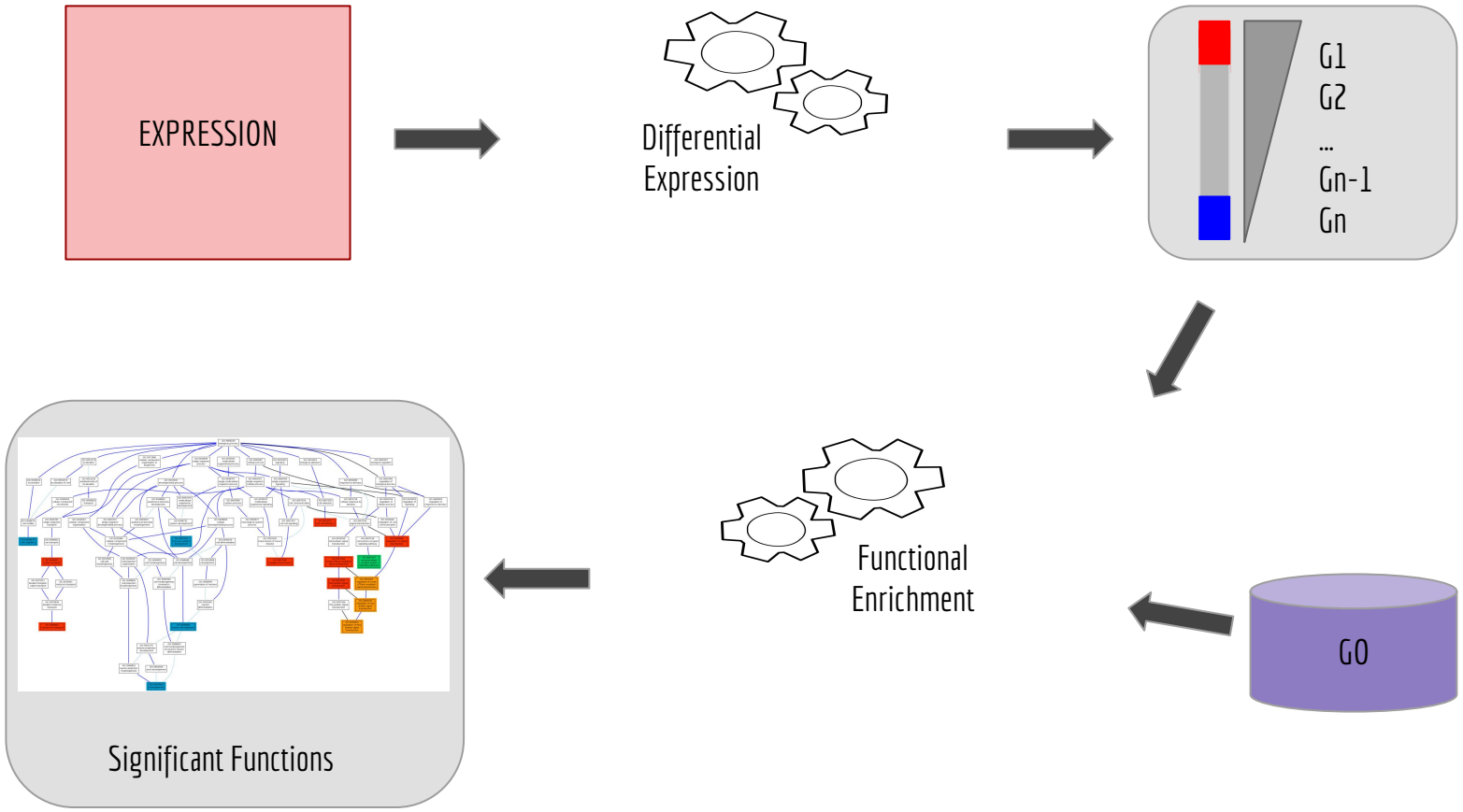
José Carbonell Caballero

Contents

- Previous steps (classic approach)
- Networks are around us
- Networks in molecular biology
- Tools
  - Network enrichment (SNOW)
    - Some exercises
  - Gene set network enrichment (Network miner)
    - Some exercises

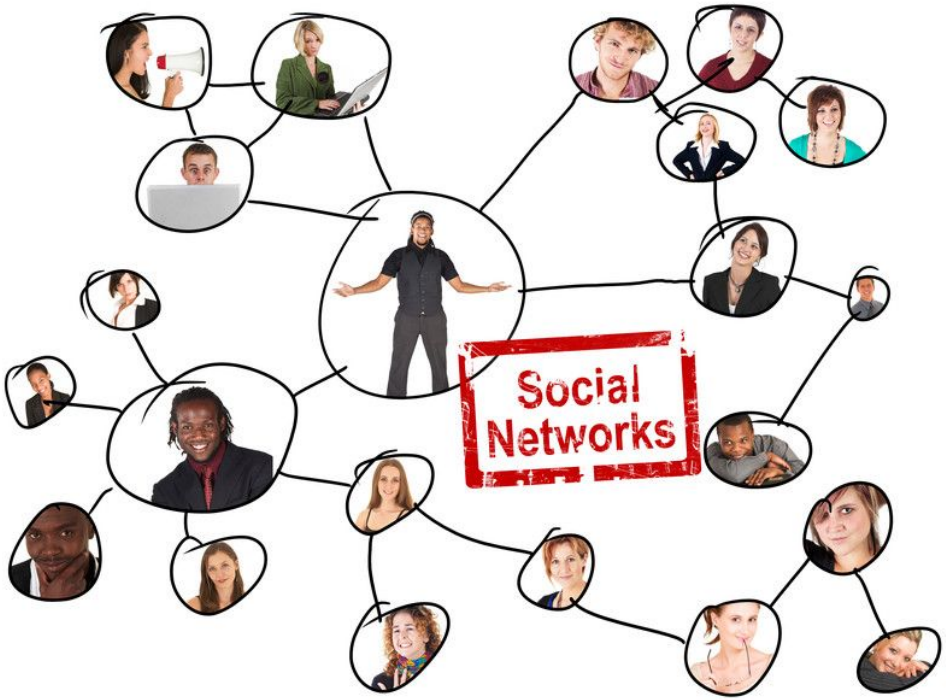
Previous steps

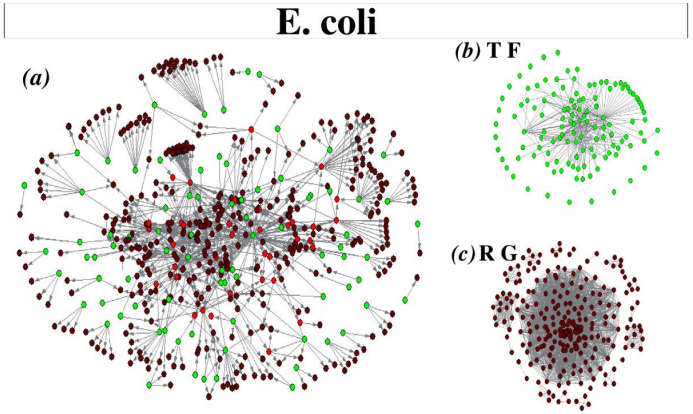
- Classic approach



Networks are around us

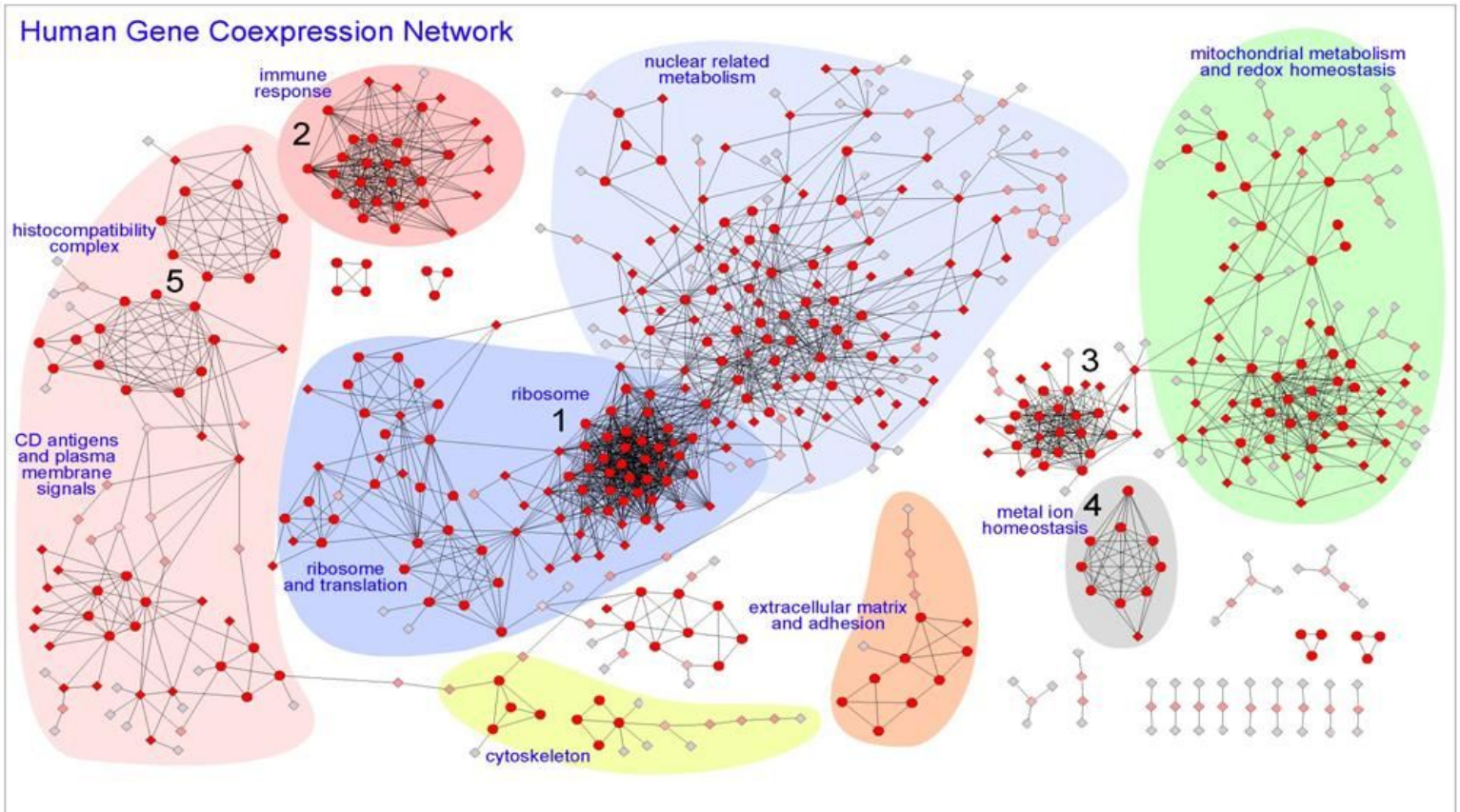
- Networks are around us





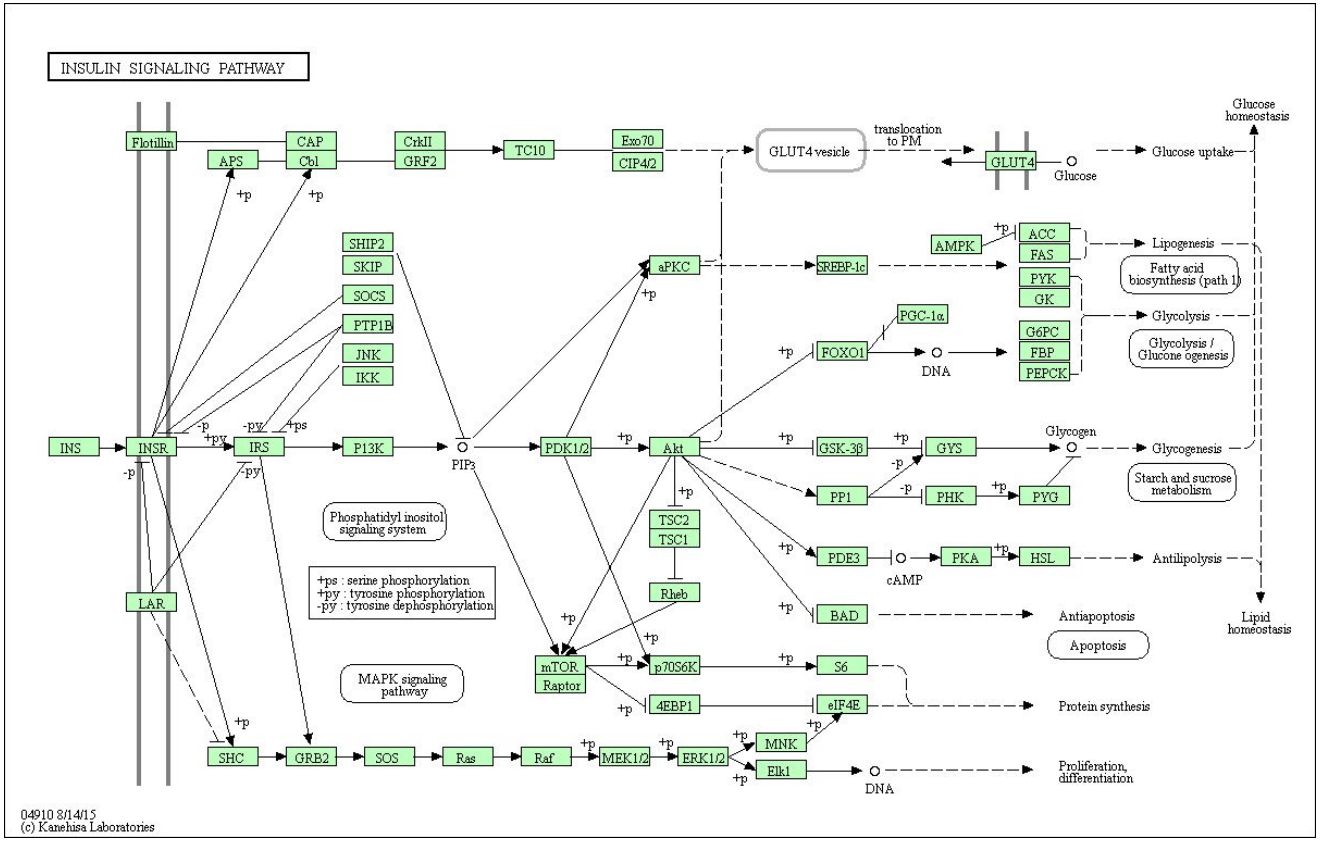
Networks are around us

- Networks are extremely useful to represent inner interactions elements within a system



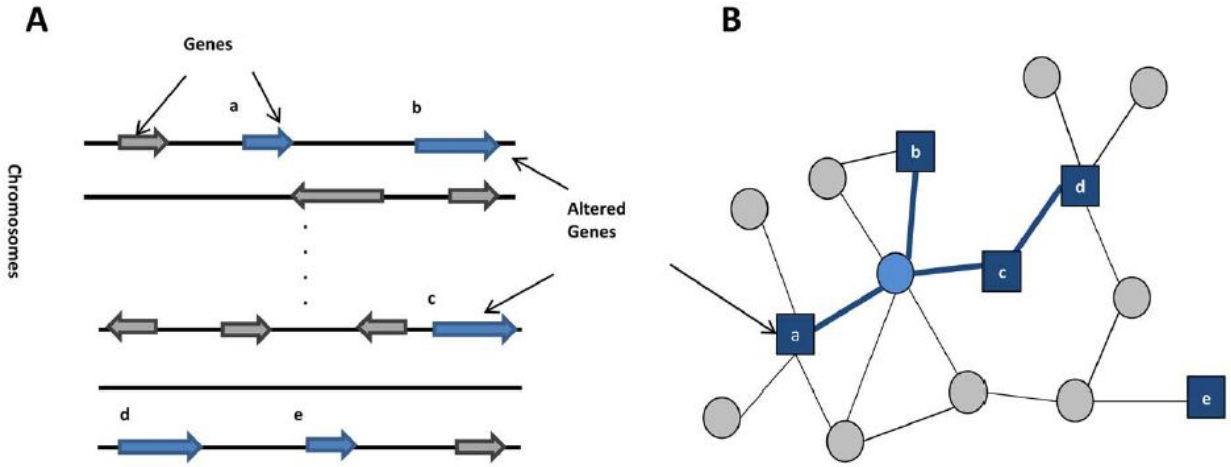
Networks in molecular biology

- Networks give us a good approach to systems biology



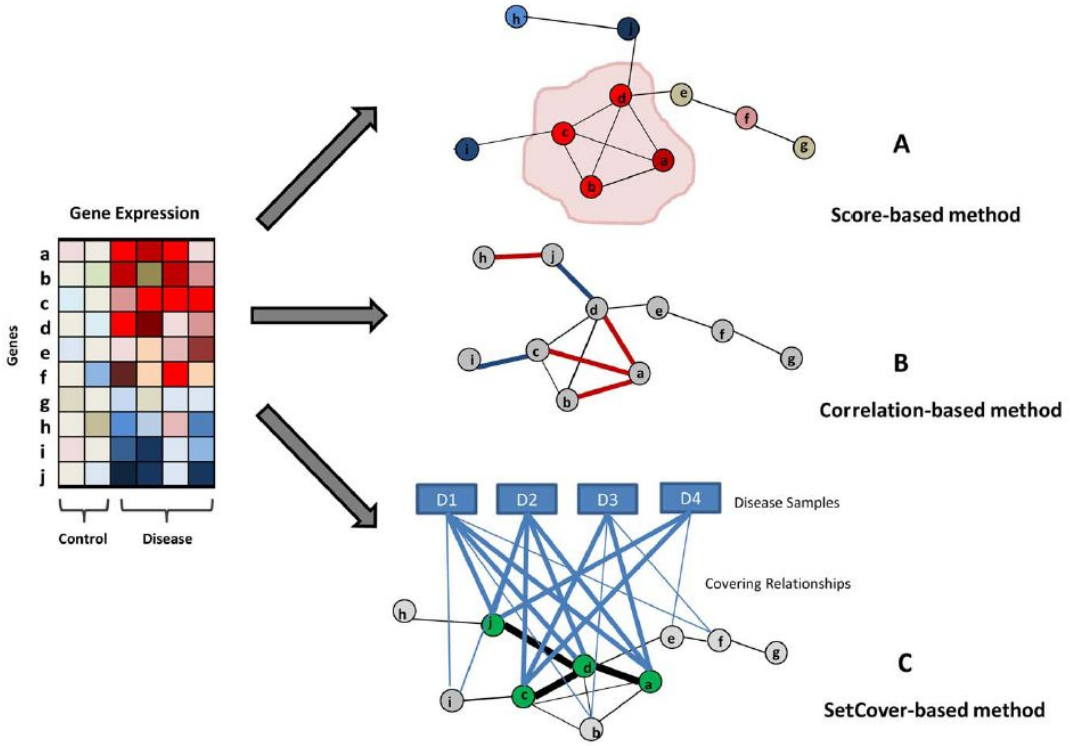
Networks in molecular biology

- Some derived concepts
  - How the system is impaired (rather than its elements)
  - Sample heterogeneity



Networks in molecular biology

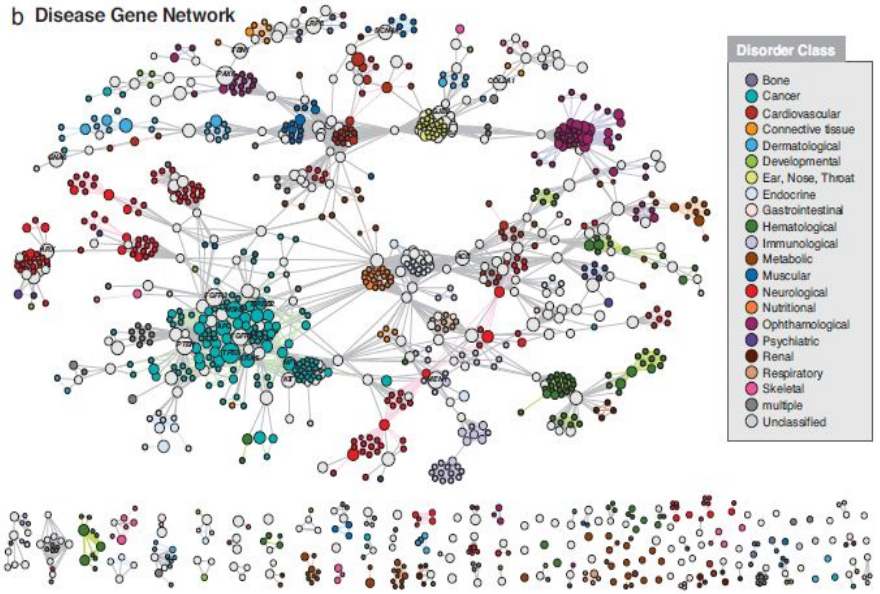
- Gene neighbourhoods (communities)
- Secondary candidates



Networks in molecular biology

Modular nature of human genetic diseases

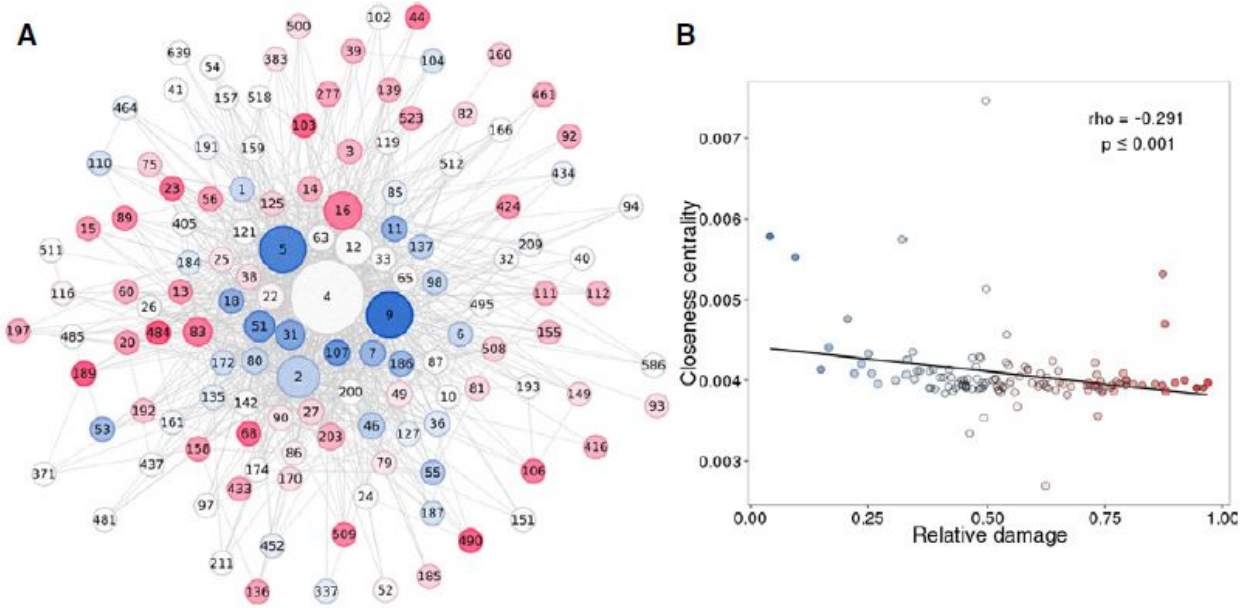
- Some complex diseases shown a high degree of sample heterogeneity (e.g. Retinitis pigmentosa)
- Same disease in different populations is caused by different genes (Fernandez, 2013, Orphanet J Rare Dis)
- Phenotypically similar diseases are often caused by functionally related genes (Goh 2007 PNAS)



Disease is better explained as a **System Perturbation** rather than a sub-cellular characteristic

Networks in molecular biology

- Structure of protein-protein interactome explains the human mutational burden in healthy population



Networks in molecular biology

Disease-gene prioritization approaches

ARTICLE

Walking the Interactome for Prioritization of Candidate Disease Genes

Sebastian Köhler,[1,2] Sebastian Bauer,[1,2] Denise Horn,[1] and Peter N. Robinson[1,*]

The identification of genes associated with hereditary disorders has contributed to improving medical care and to a better understanding of gene functions, interactions, and pathways. However, there are well over 1500 Mendelian disorders whose molecular basis remains unknown. At present, methods such as linkage analysis can identify the chromosomal region in which unknown disease genes are located, but the regions could contain up to hundreds of candidate genes. In this work, we present a method for prioritization of candidate genes by use of a global network distance measure, random walk analysis, for definition of similarities in protein-protein interaction networks. We tested our method on 110 disease-gene families with a total of 783 genes and achieved an area under the ROC curve of up to 98% on simulated linkage intervals of 100 genes surrounding the disease gene, significantly outperforming previous methods based on local distance measures. Our results not only provide an improved tool for positional-cloning projects but also add weight to the assumption that phenotypically similar diseases are associated with disturbances of subnetworks within the larger protein interactome that extend beyond the disease proteins themselves.

BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 10. NO 5. 280–293 doi:10.1093/bfgp/elr024

Network-based methods for human disease gene prediction

Xiujuan Wang, Natali Gulbahce and Haiyuan Yu

Advance Access publication date 15 July 2011

Abstract

Despite the considerable progress in disease gene discovery, we are far from uncovering the underlying cellular mechanisms of diseases since complex traits, even many Mendelian diseases, cannot be explained by simple genotype–phenotype relationships. More recently, an increasingly accepted view is that human diseases result from perturbations of cellular systems, especially molecular networks. Genes associated with the same or similar diseases commonly reside in the same neighborhood of molecular networks. Such observations have built the basis for a large collection of computational approaches to find previously unknown genes associated with certain diseases. The majority of the methods are based on protein interactome networks, with integration of other large-scale genomic data or disease phenotype information, to infer how likely it is that a gene is associated with a disease. Here, we review recent, state of the art, network-based methods used for prioritizing disease genes as well as unraveling the molecular basis of human diseases.

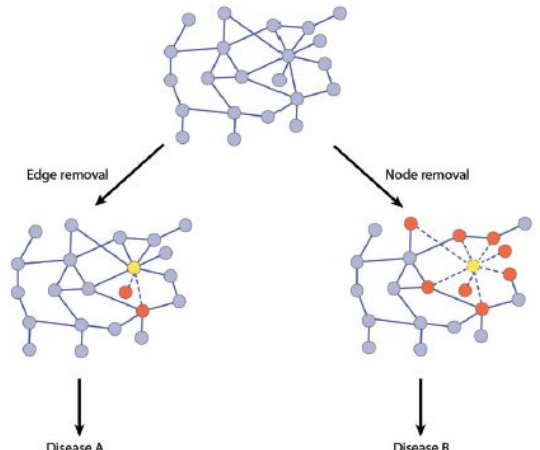
Keywords: *human diseases; disease network; disease gene prediction; protein–protein interaction; molecular network*

Bioinformatics Advance Access published July 30, 2014

Walking the interactome for candidate prioritization in exome sequencing studies of Mendelian diseases

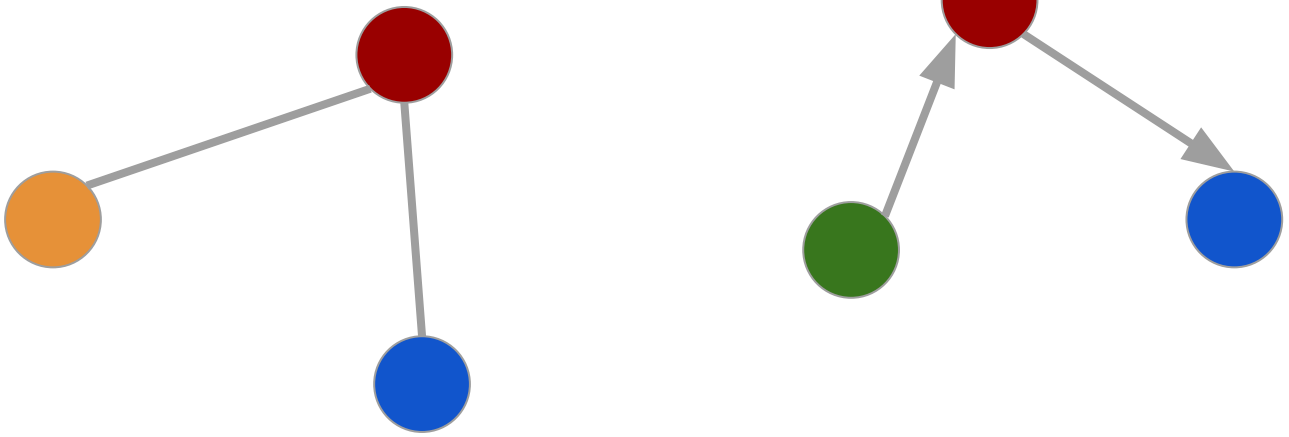
Damian Smedley [1,†], Sebastian Köhler [2,†], Johanna Christina Czeschik [3], Joanna Amberger [4], Carol Bocchini [4], Ada Hamosh [4], Julian Veldboer [2,5], Tomasz Zemojtel [2,6], and Peter N Robinson [2,5,7,8*]

[1]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK, [2]Institute for Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany, [3]Genome Informatics, Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Hufelandstr. 55, 45122 Essen, Germany, [4] Baltimore, MD, USA, [5]Institute for Bioinformatics, Department of Mathematics and Computer Science, Freie Universität Berlin, Takustr. 9, 14195 Berlin, [6]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland [7]Berlin-Brandenburg Center for Regenerative Therapies, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany, [8]Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany



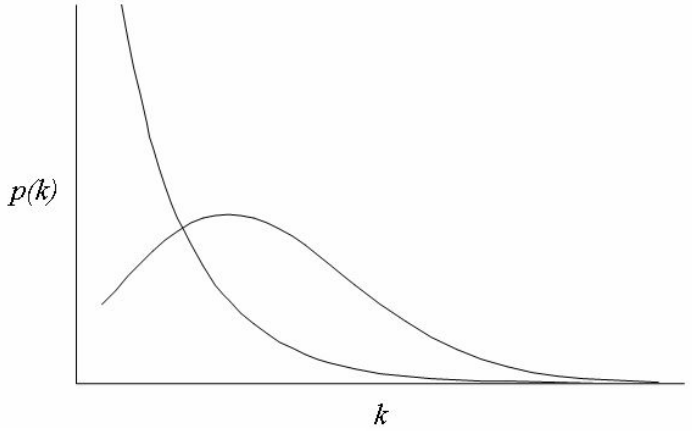
Networks in molecular biology

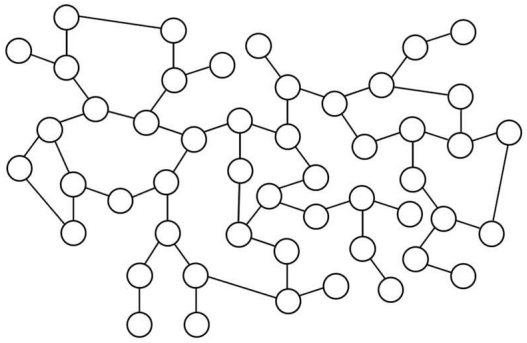
- Knowledge is represented as a graph
  - Proteins are the vertices
  - Interactions are the edges



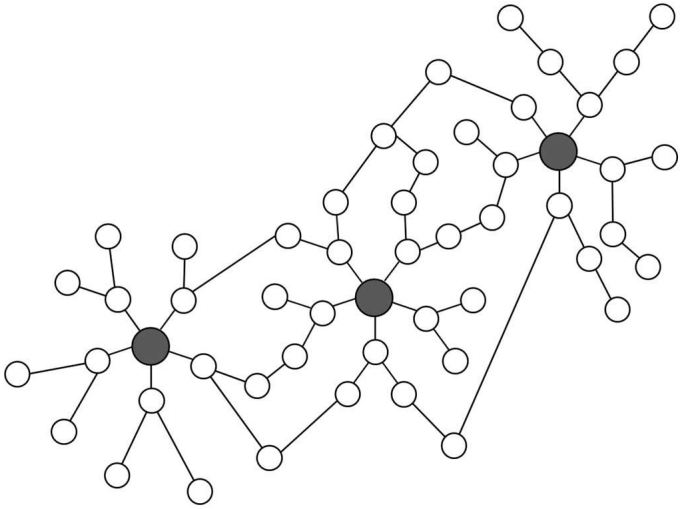
Networks in molecular biology

- Biological networks
  - Scale-free (power law)





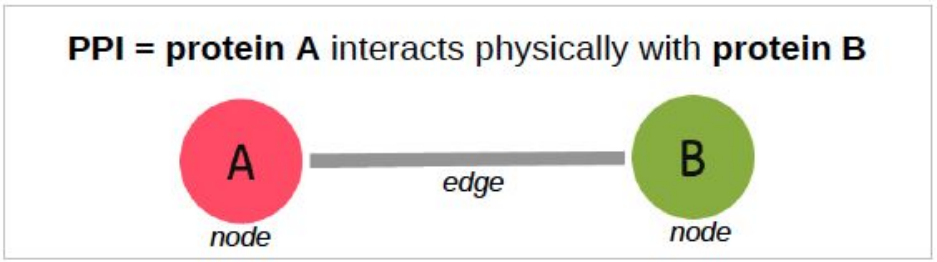
(a) Random network

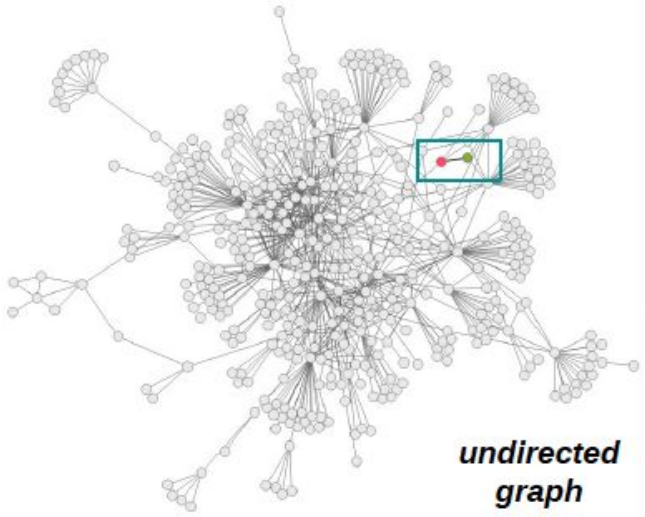


(b) Scale-free network

Networks in molecular biology

- Types of networks
  - Protein-protein interactions

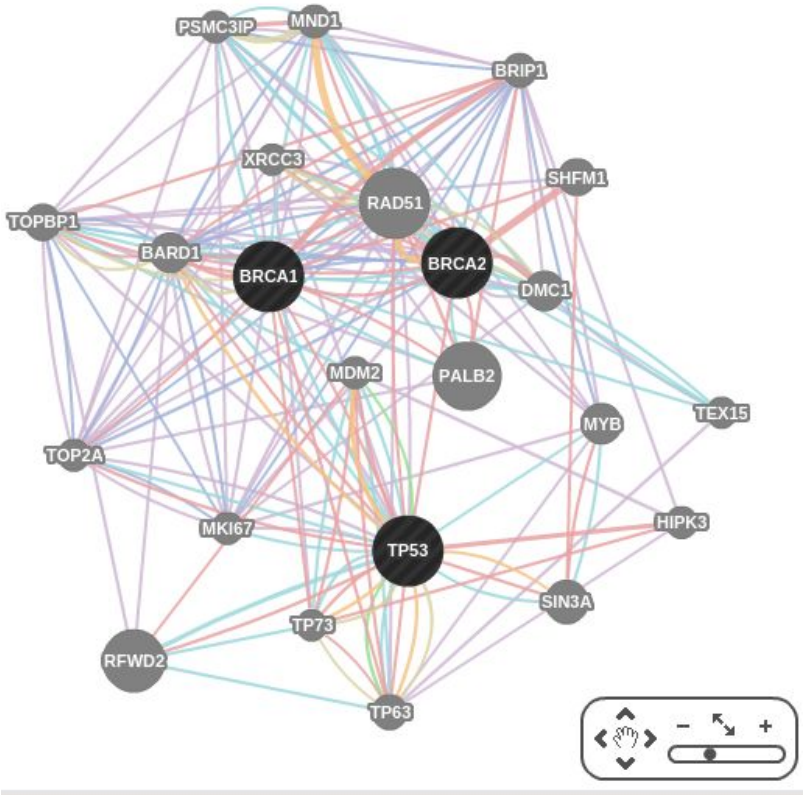




Networks in molecular biology

- Types of networks
  - PTMs
  - Coexpression
  - Functional terms
  - Mixed networks

Physical interactions	64.66 %
Co-expression	17.38 %
Predicted	7.17 %
Pathway	5.04 %
Co-localization	3.22 %
Genetic interactions	1.68 %
Shared protein domains	0.84 %

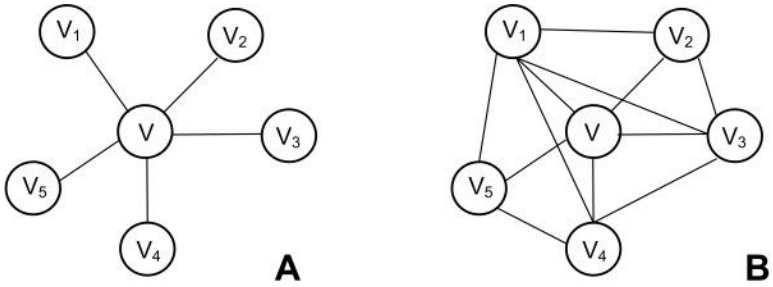


http://genemania.org/

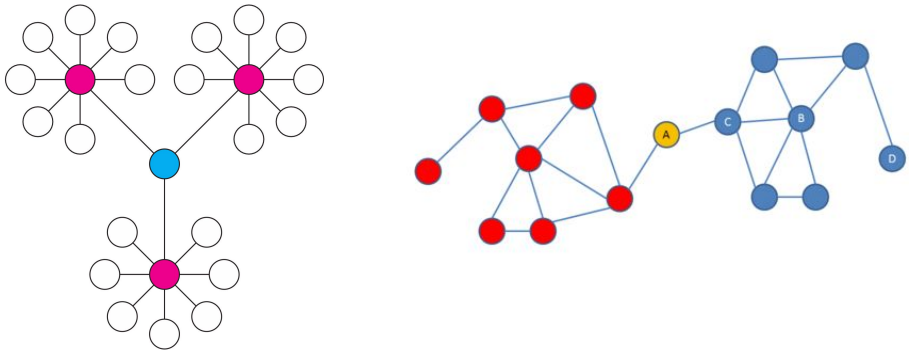
Networks in molecular biology

Network are evaluated through some parameters:

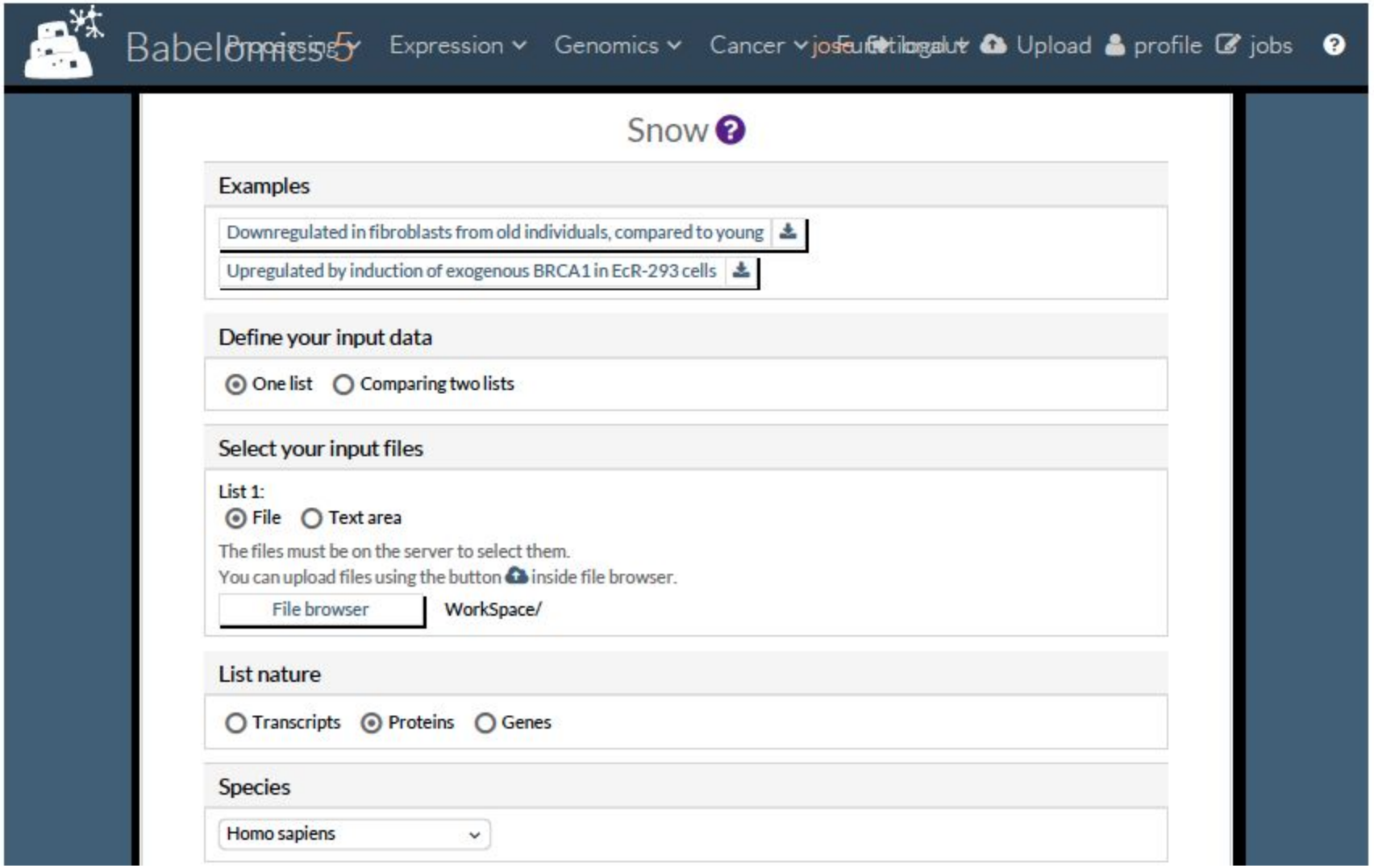
- At a network level:
  - Number of components
  - Clustering coefficient



- At a gene level:
  - Closeness centrality
  - Betweenness centrality
  - Degree

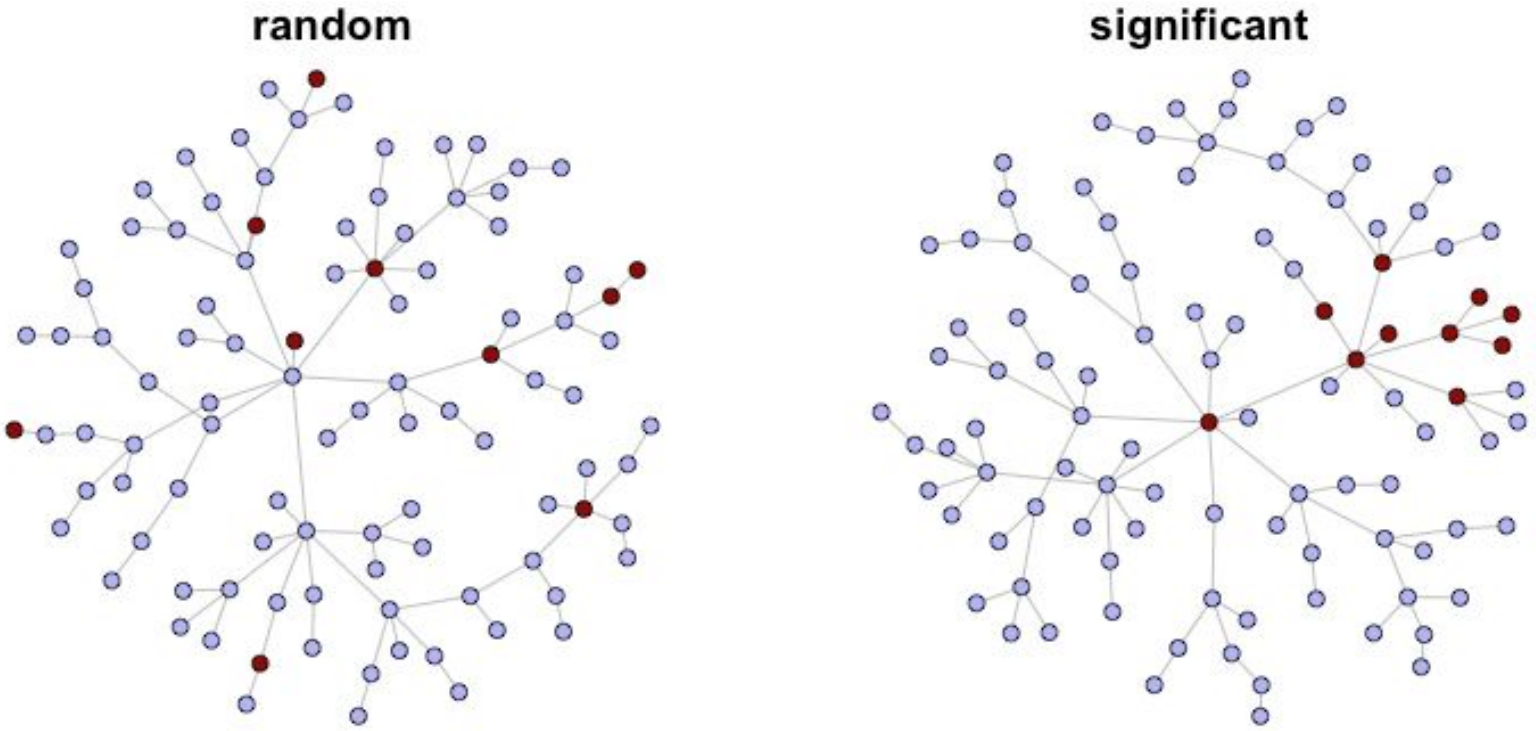


Network enrichment (SNOW)

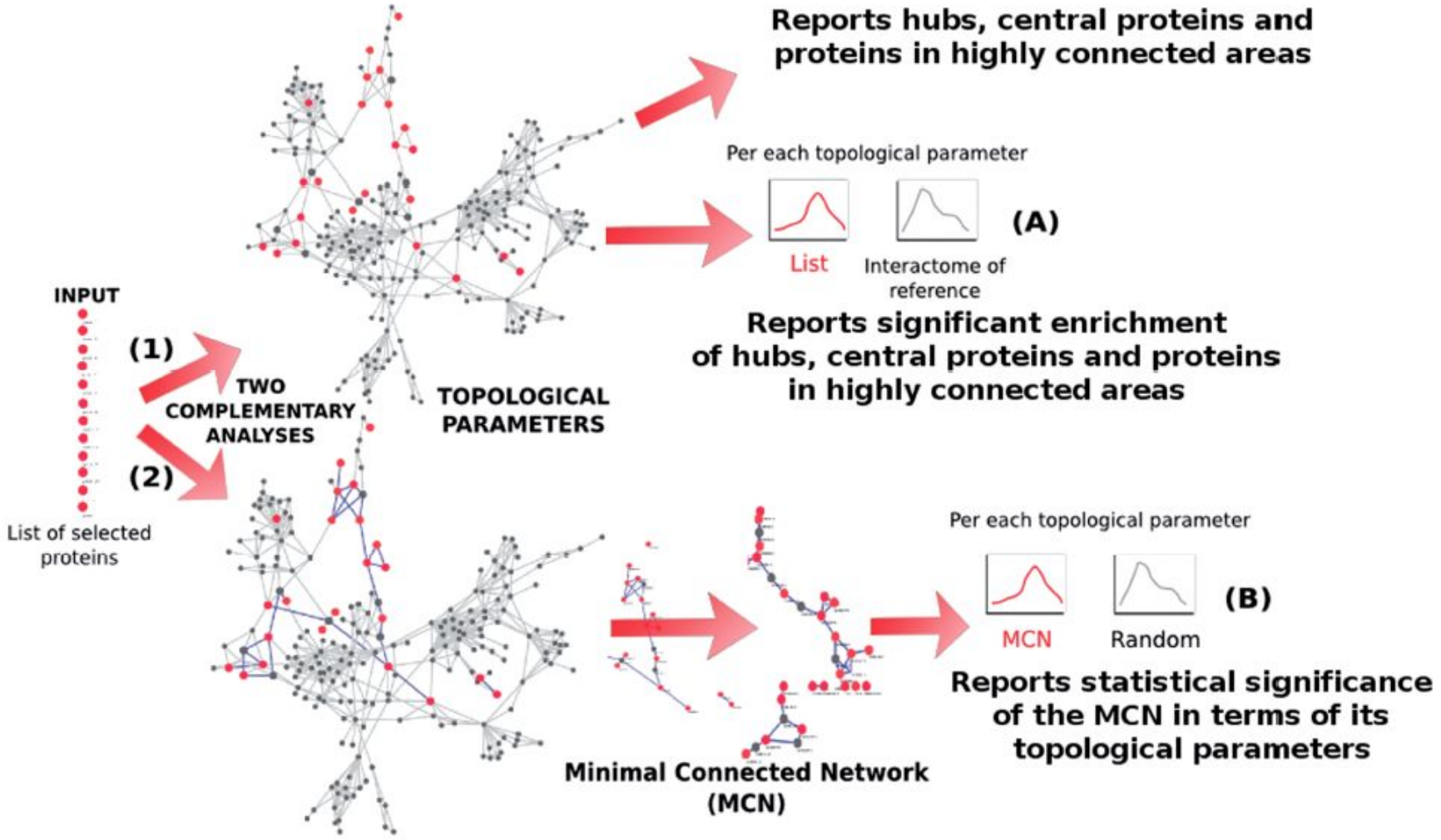


Network enrichment (SNOW)

Does a set of input genes represent a biologically meaningful network?

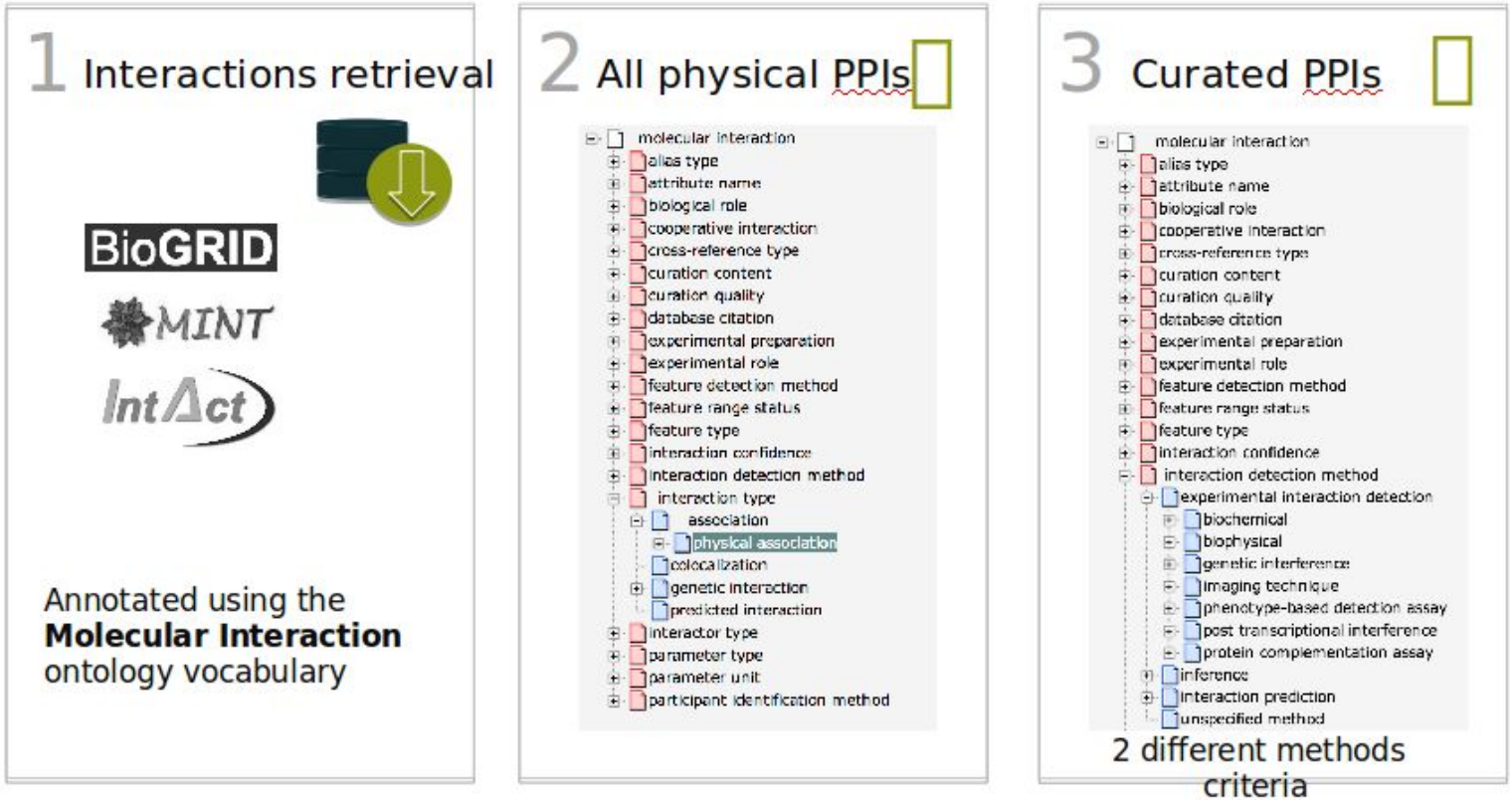


Network enrichment (SNOW)

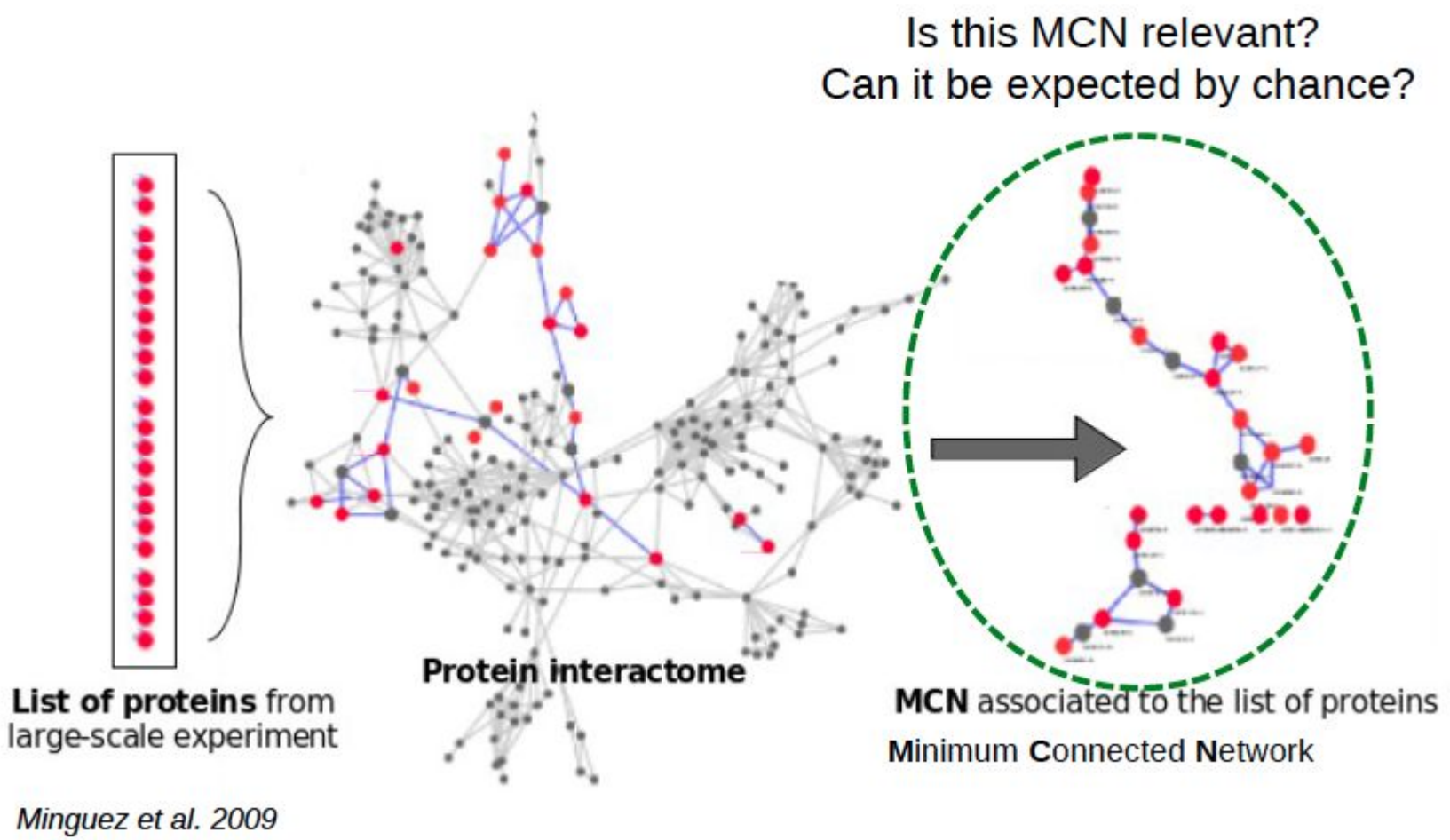


Network enrichment (SNOW)

Network compilation



Network enrichment (SNOW)

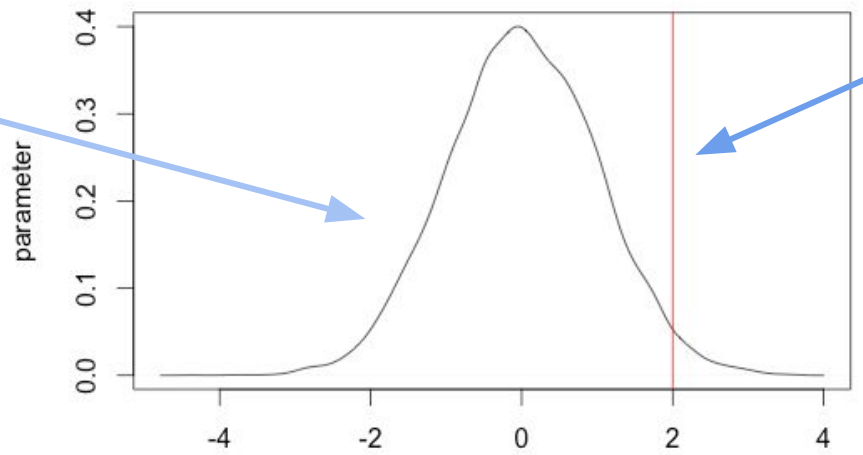


Minguez et al. 2009

Network enrichment (SNOW)

- Comparison against N random networks

Random networks (NULL distribution)

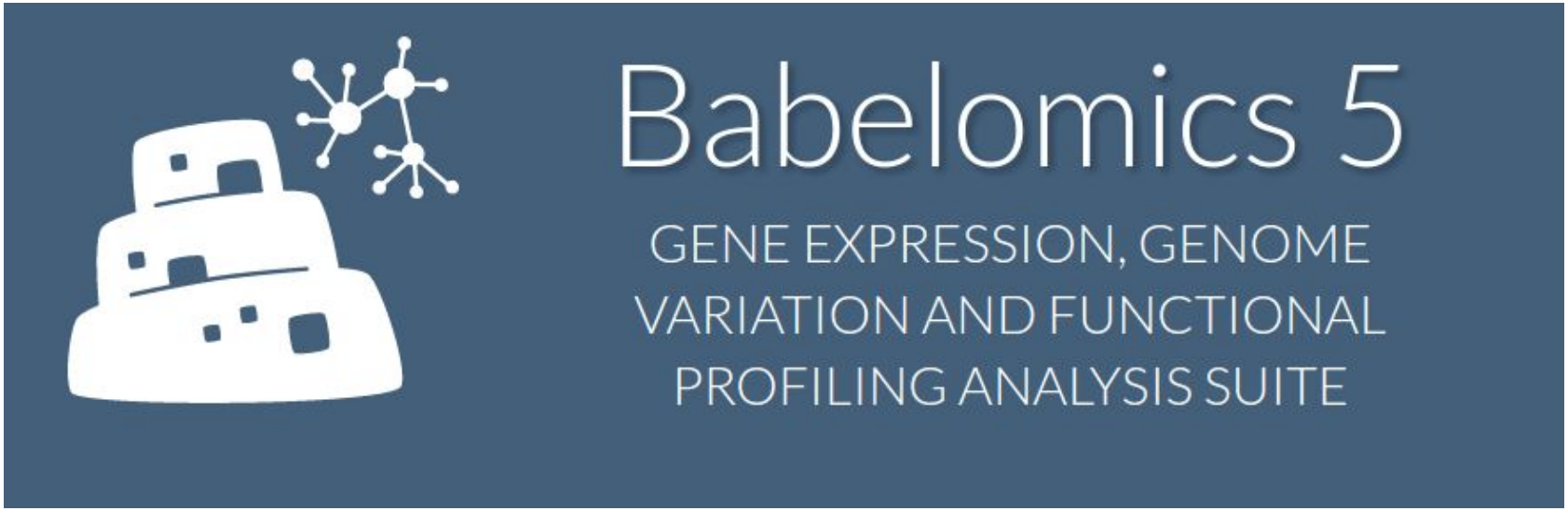


Input network

Some exercices

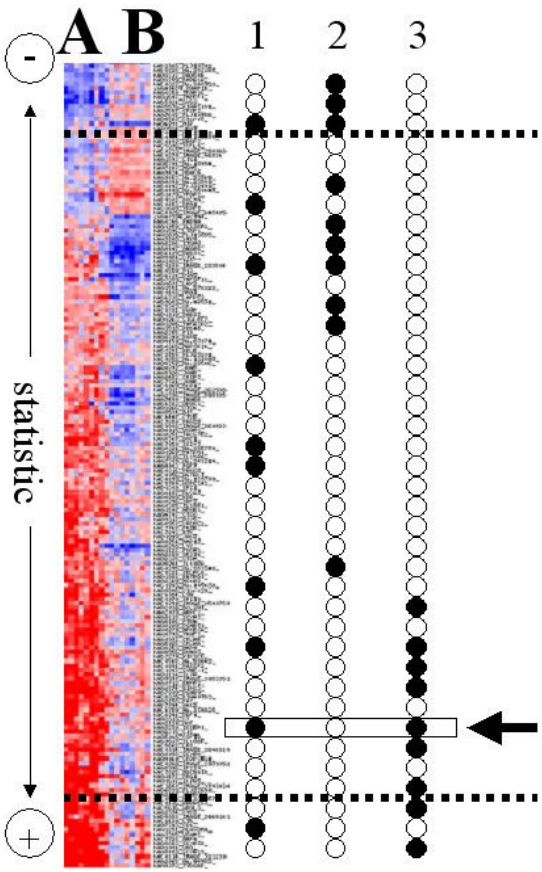
Worked examples of SNOW in Babelomics 5 wiki

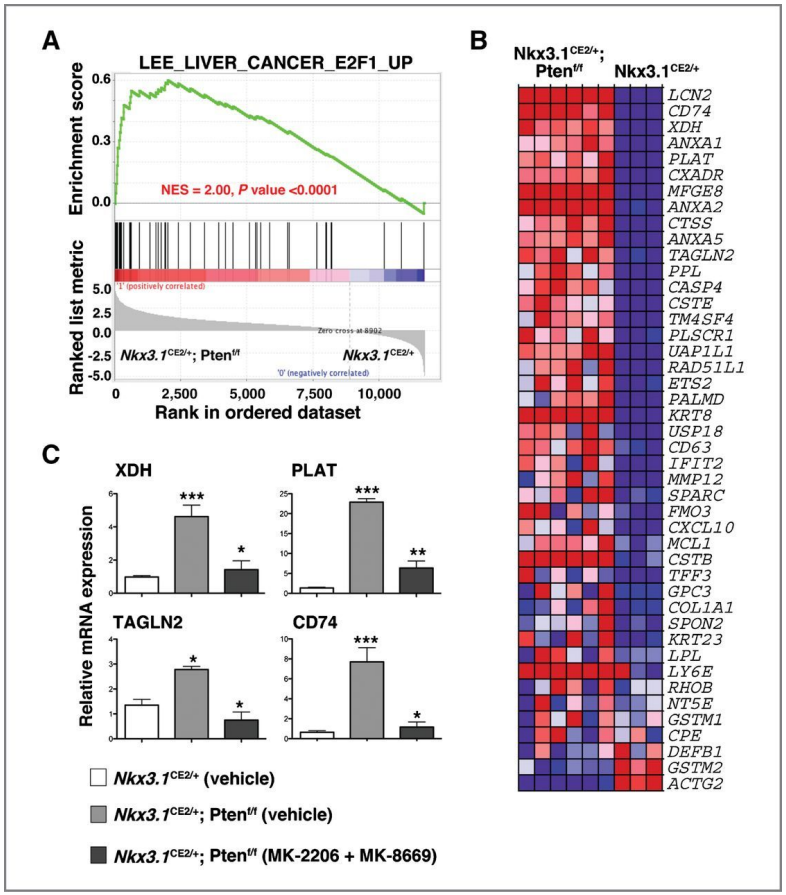
https://github.com/babelomics/babelomics/wiki/network-enrichment-(snow)



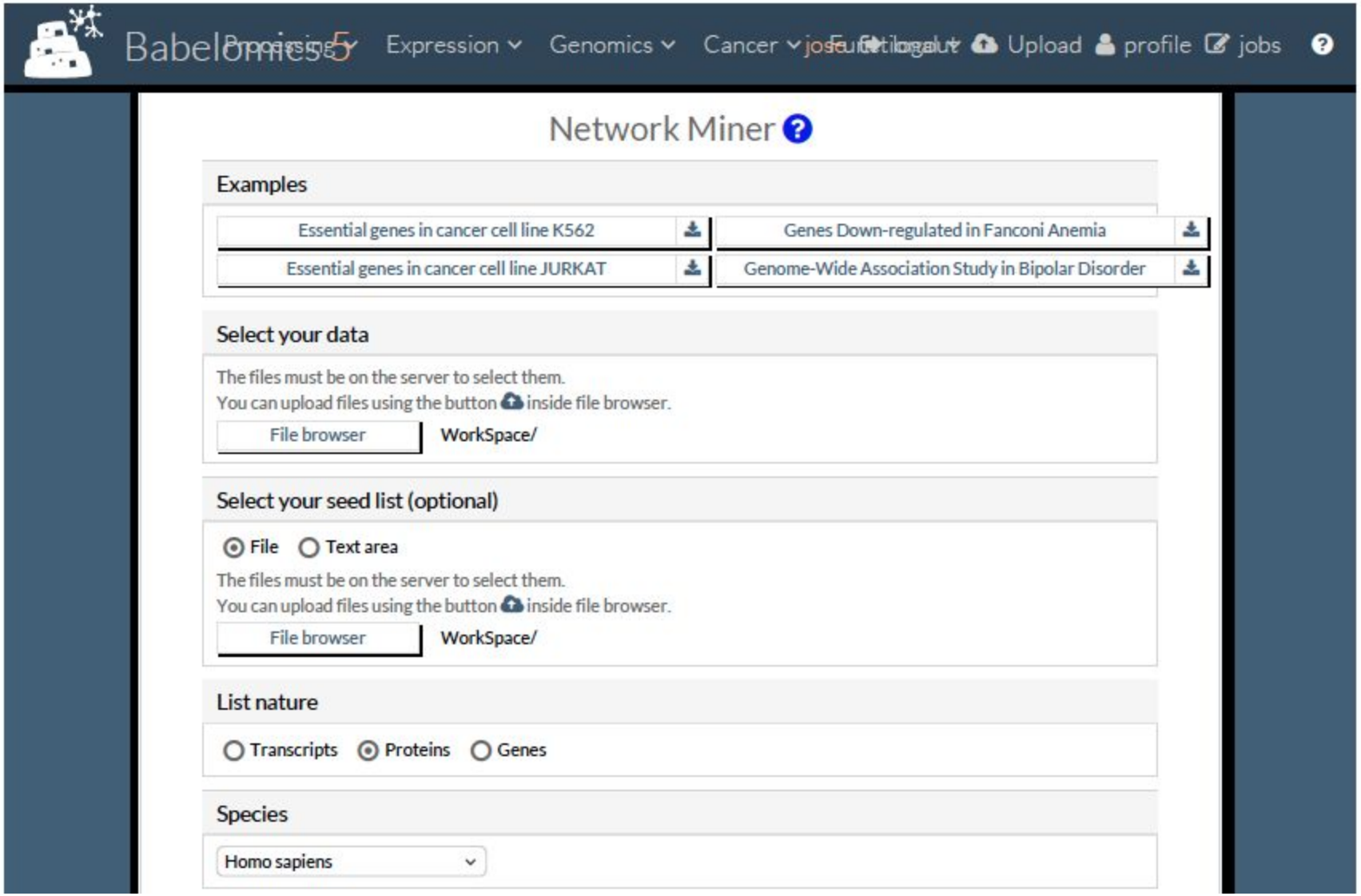
Network-based GSEA (Network miner)

- Gene set enrichment



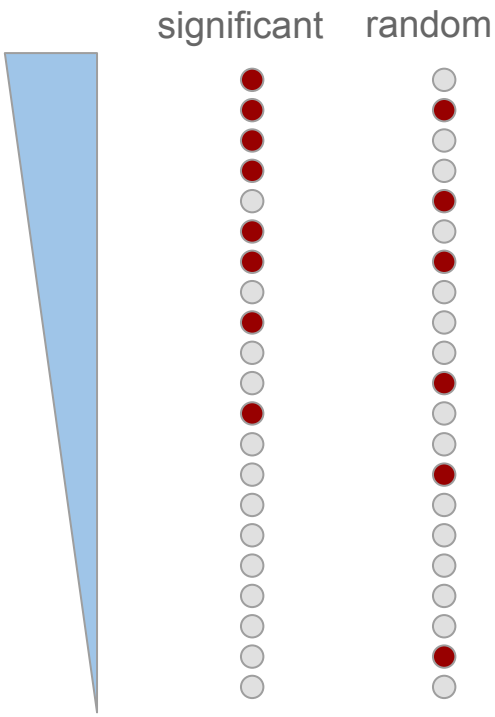


Network-based GSEA (Network miner)

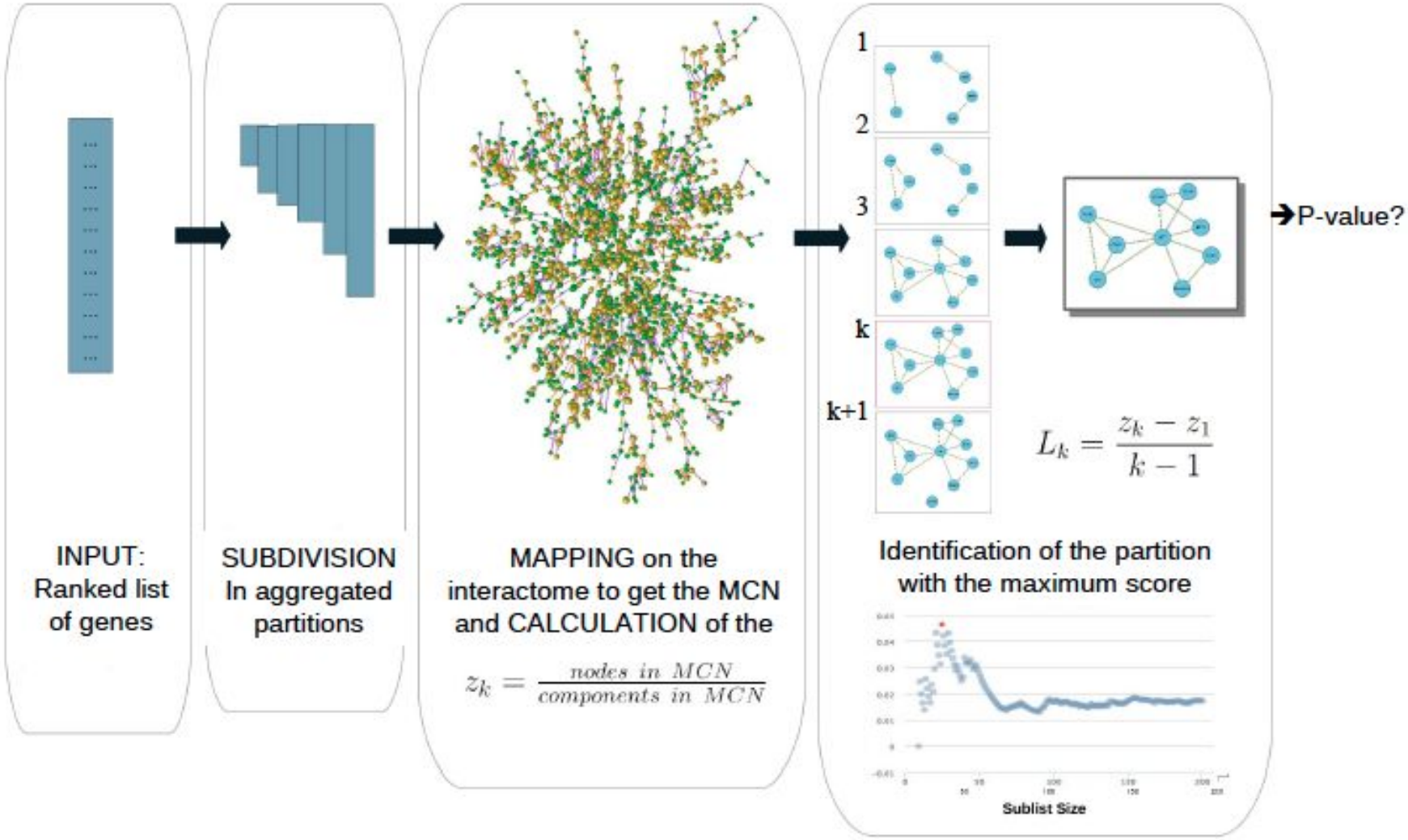


Network-based GSEA (Network miner)

Is there a latent sub-network related to ranking criteria?



Network-based GSEA (Network miner)



Network-based GSEA (Network miner)

Enrichment score

(a) First, ordering the parameter of interest z_k according to the ranked list, all relative maxima are identified. The partitions so selected (S_k^{max}) represent situations where a new protein capable of connecting to the previous ones is added to the previous partitions.

(b) Second, the score L_k *is* computed as $L_k = (z_k - 1)/(k-1)$ for all the selected partitions S_k^{max}. The score can be seen as a balance between the increase in connected nodes and the distance to the top of the ranked list ($k = 1$).

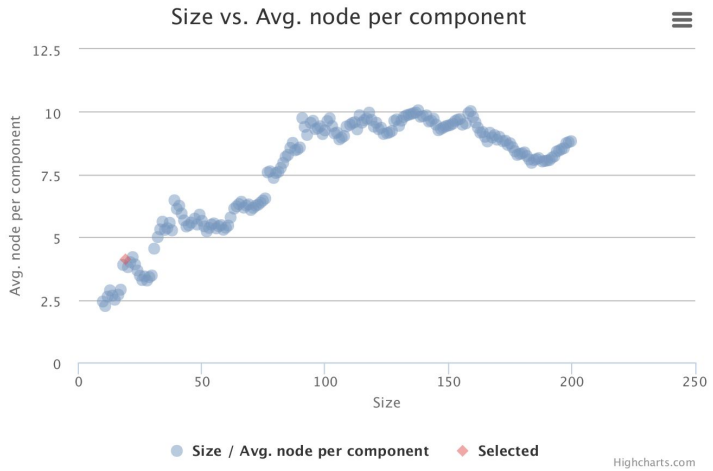
(c) Third, we choose the partition S_{best} and index k_{best} corresponding to the highest L_k computed in b) form the S_k^{max} chosen in (a).

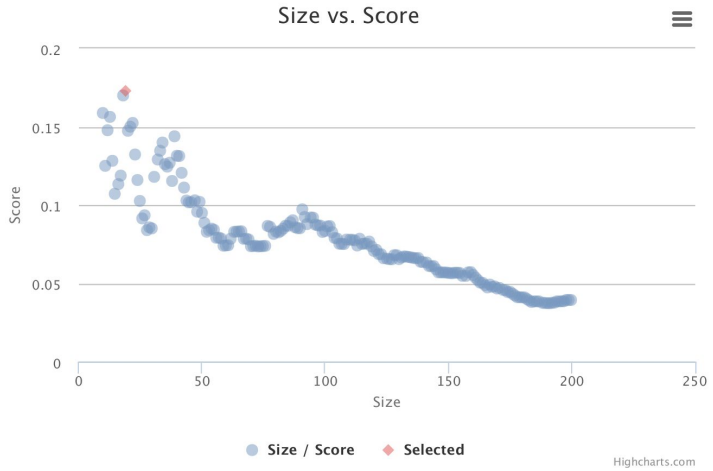
Results: Minimum Connected Network selected

All results

MCN selected pval: *<0.0035*

MCN selected size: *19*



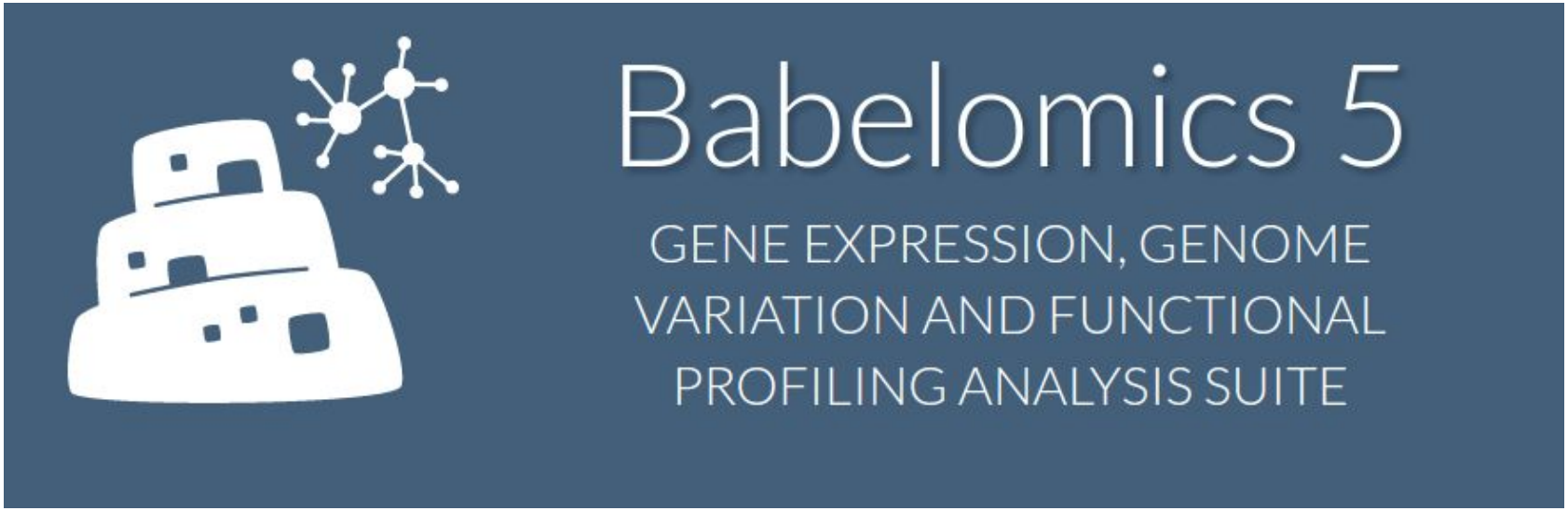


All results *result_all.txt*

Some exercises…

Worked examples of NM on Babelomics 5 wiki

https://github.com/babelomics/babelomics/wiki/Gene%20Set%20Network%20Enrichment%20(Network%20Miner)



Some conclusions

- Networks are extremely useful to represent complex interactions between the components of a living system
- Networks (derived from selected genes) can be characterized by using different topological parameters.
- Network-based representations give us a perfect approach for systems biology
- Mechanistic interpretation of impaired elements is quite intuitive (at least compared to f.e.)

References

- Minguez, P., Götz, S., Montaner, D., Al-Shahrour, F., & Dopazo, J. (2009). SNOW, a web-based tool for the statistical analysis of protein-protein interaction networks. *Nucleic Acids Research*, *37*(SUPPL. 2), 109–114. doi:10.1093/nar/gkp402
- García-Alonso, L., Alonso, R., Vidal, E., Amadoz, A., De María, A., Minguez, P., … Dopazo, J. (2012). Discovering the hidden sub-network component in a ranked list of genes or proteins derived from genomic experiments. *Nucleic Acids Research*, *40*(20), 1–13. doi:10.1093/nar/gks699
- Garcia-alonso, L., Jiménez-almazán, J., Carbonell-caballero, J., Vela-boza, A., Santoyo-lópez, J., Antiñolo, G., & Dopazo, J. (2014). The role of the interactome in the maintenance of deleterious variability in human populations. Molecular Systems Biology, 1–18.
- Prieto, C., Risueño, A., Fontanillo, C., & De Las Rivas, J. (2008). Human gene coexpression landscape: Confident network derived from tissue transcriptomic profiles. PLoS ONE, 3(12). doi:10.1371/journal.pone.0003911
- Albert, R., Jeong, H., & Barabasi, A. (2000). Error and attack tolerance of complex networks. Nature, 406(6794), 378–82. doi:10.1038/35019019
- Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., … Morris, Q. (2010). The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. Nucleic Acids Research, 38(SUPPL. 2), 214–220. doi:10.1093/nar/gkq537
- Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. Proceedings of the National Academy of Sciences of the United States of America, 104(21), 8685–8690. doi:10.1073/pnas.0701361104
- Cho, D., Kim, Y., & Przytycka, T. M. (2012). Chapter 5: Network biology approach to complex diseases. PLoS Computational Biology, 8(12), e1002820. doi:10.1371/journal.pcbi.1002820

Thank you

Any comment/question is welcome

jcarbonell@cipf.es