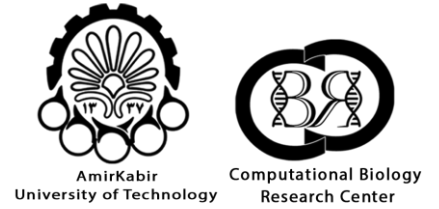




Morteza Kouhsar



MSC: Computer Science, Tarbiat Modares University.

PhD: Bioinformatics, Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics (IBB), University of Tehran.



Kouhsar, Morteza, Fatemeh Zare-Mirakabad, and Yousef Jamali. "WCOACH: Protein complex prediction in weighted PPI networks." *Genes & genetic systems* 90, no. 5 (2015): 317-324.

Motieghader, Habib, **Morteza Kouhsar**, Ali Najafi, Balal Sadeghi, and Ali Masoudi-Nejad. "mRNA-miRNA bipartite network reconstruction to predict prognostic module biomarkers in colorectal cancer stage differentiation." *Molecular BioSystems* 13, no. 10 (2017): 2168-2180.

Mehranfar, Adele, Nasser Ghadiri, **Morteza Kouhsar**, and Ashkan Golshani. "A Type-2 fuzzy data fusion approach for building reliable weighted protein interaction networks with application in protein complex detection." *Computers in Biology and Medicine* 88 (2017): 18-31.

Computational Methods on a System Level: PPI Networks



Morteza Kouhsar

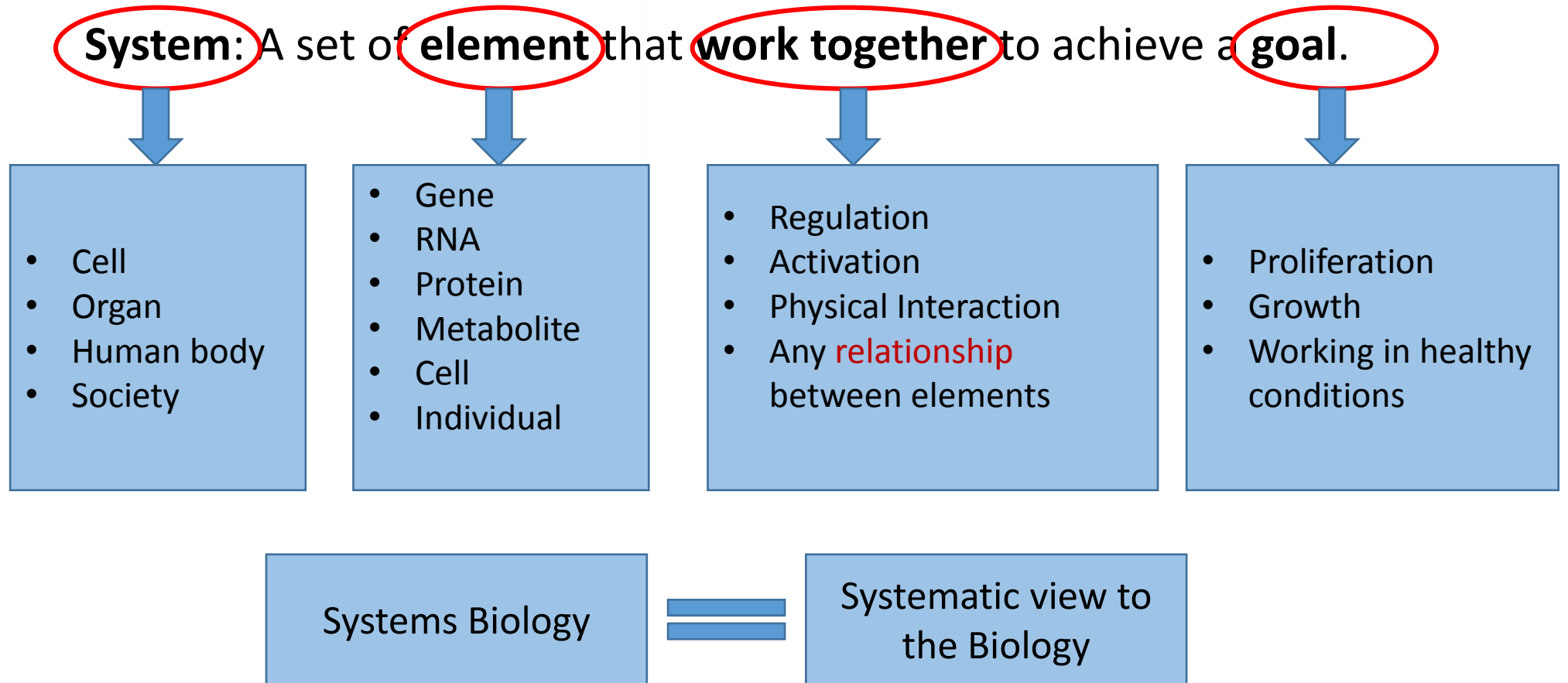
Laboratory of Systems Biology and Bioinformatics (LBB)

Outline

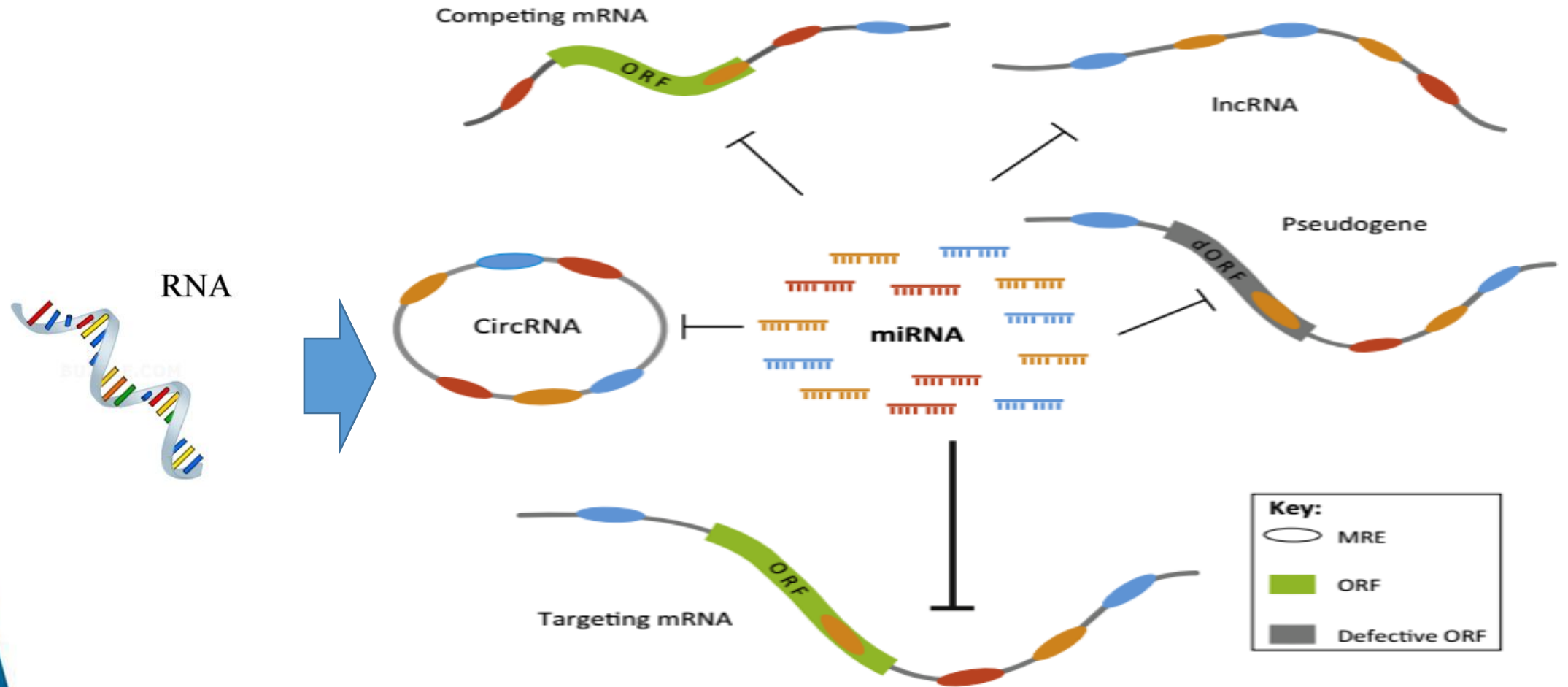
- Systems Biology and Biological Networks
- PPI Networks
 - Protein Physical Interactions
 - Resources and Databases
- PPI Network Analysis
 - Module Analysis
 - Module Detection
 - Gene Ontology and Enrichment Analysis
 - Centrality Measures and Node Analysis
- Cytoscape: A Powerful Framework for Systems Biology Analysis
 - Import Data
 - Installing Apps
 - Module detection and Enrichment



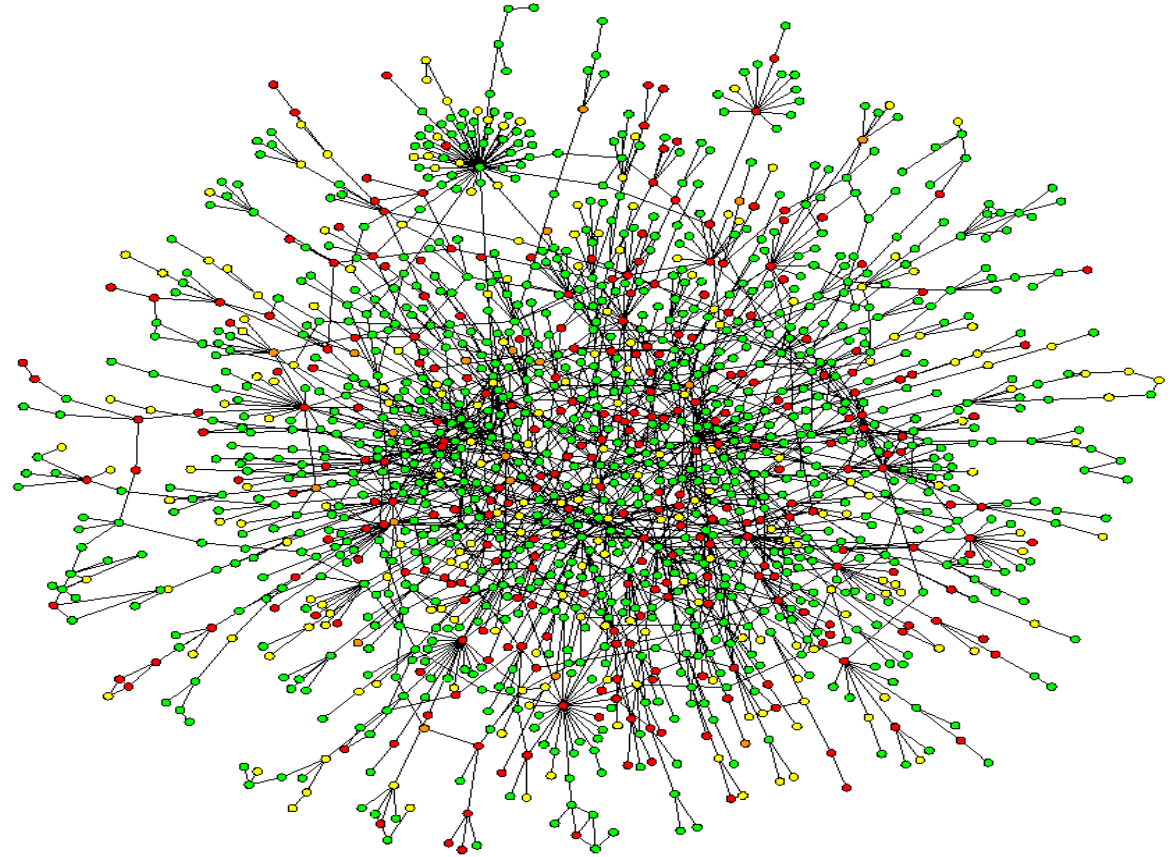
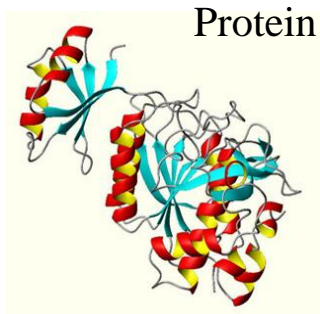
Systems Biology



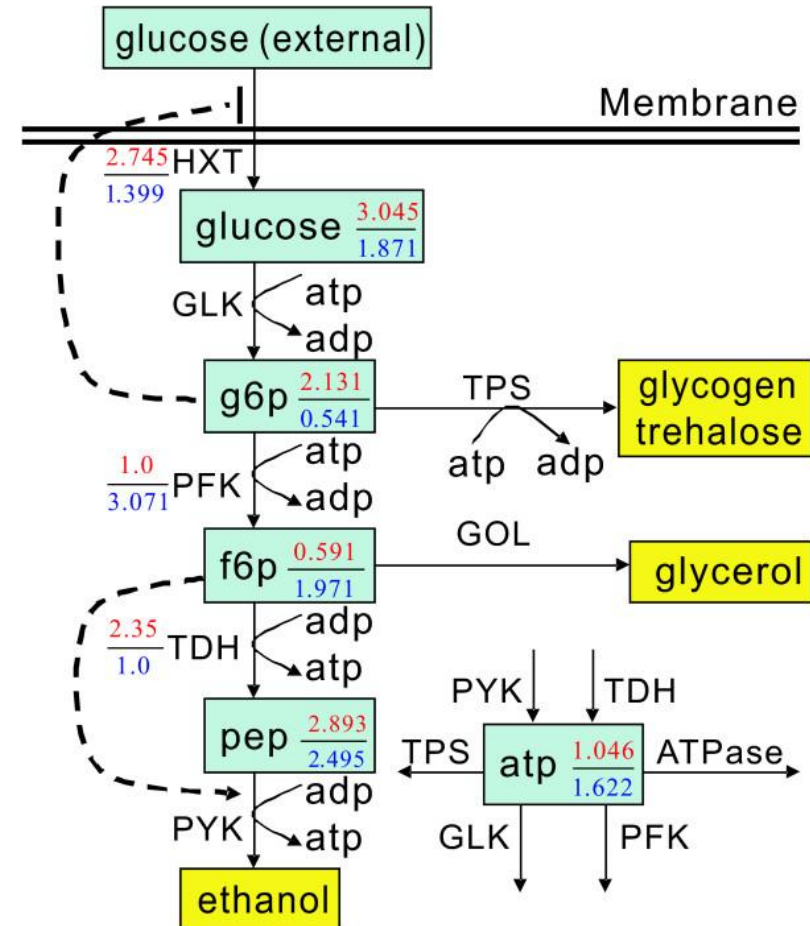
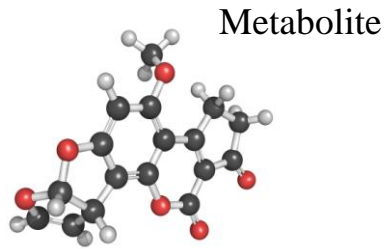
Systems Biology



Systems Biology

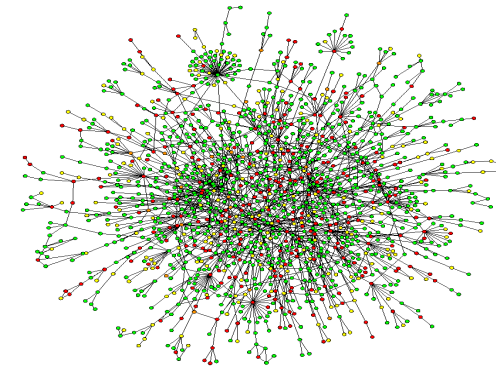
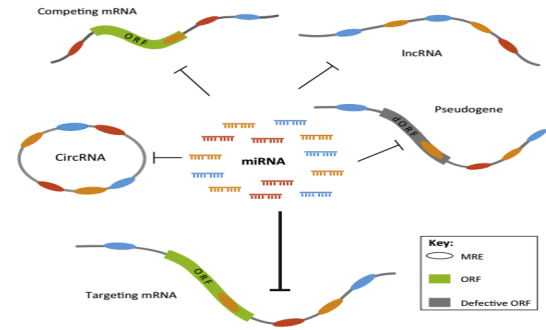
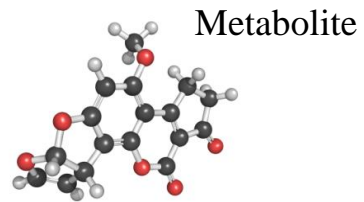
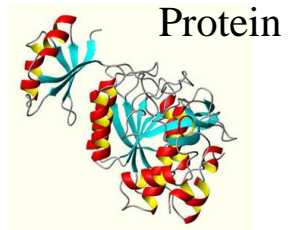
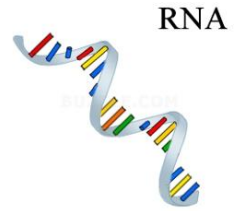


Systems Biology

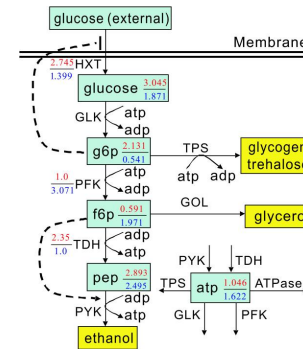


Systems Biology

Biological
Elements



Biological
Systems or
Networks



1st International Computational Biology workshop

Systems Biology

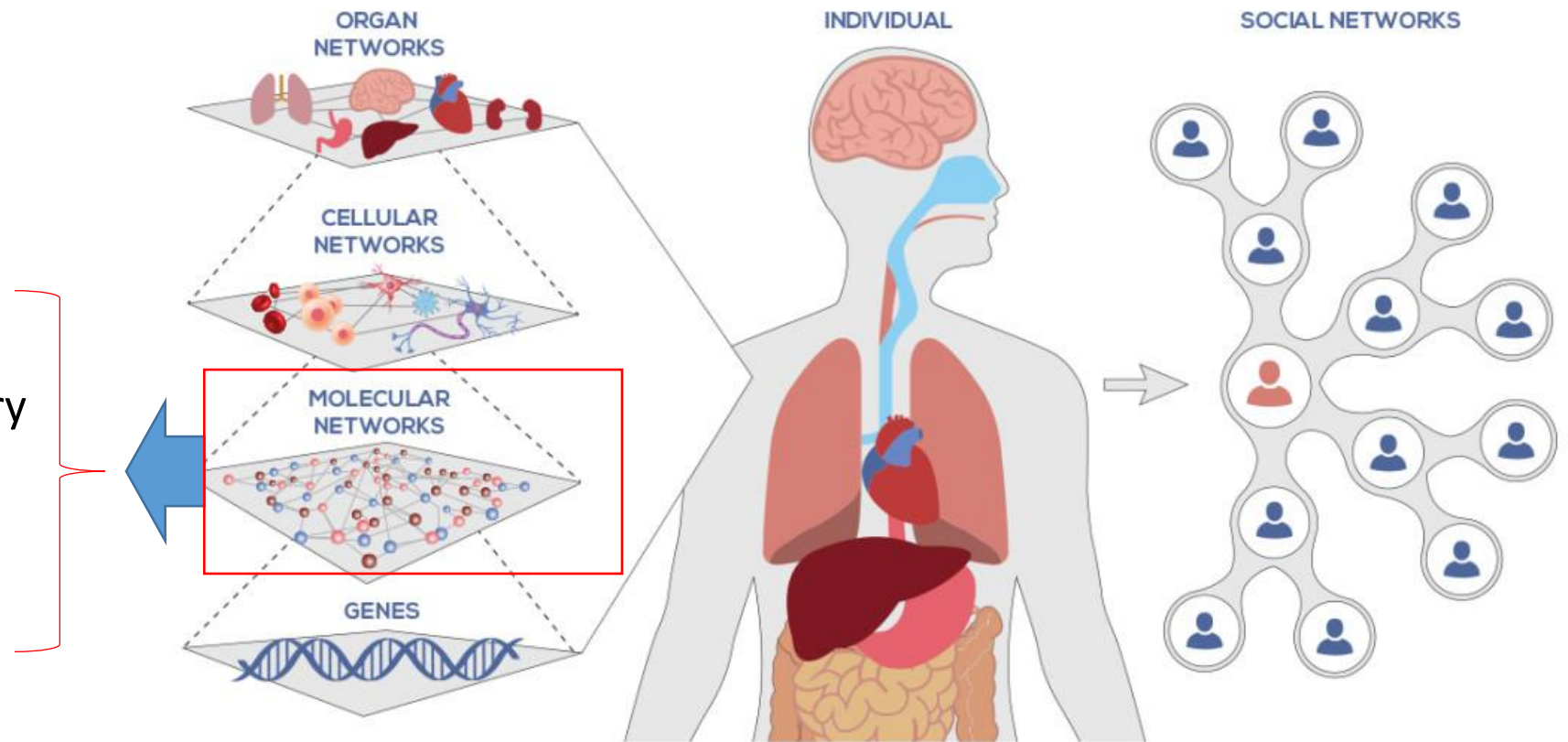
Analysis of **Biological Systems**
instead of
Biological elements individually
to understand
System behavior



Systems Biology Resources

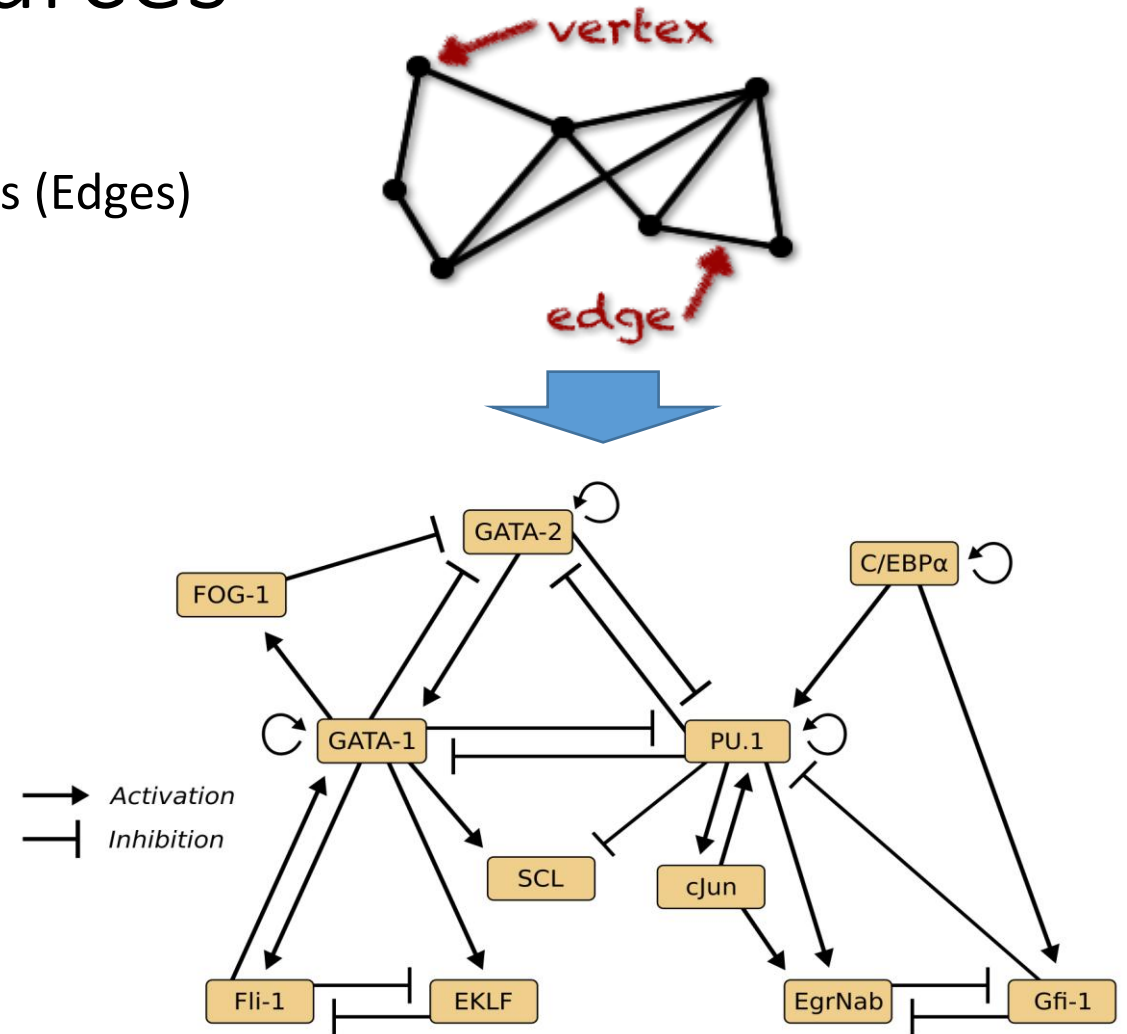
- Biological Networks:

- PPI
- Gene Regulatory
- Metabolic
- DNA-Protein
- Co-expression



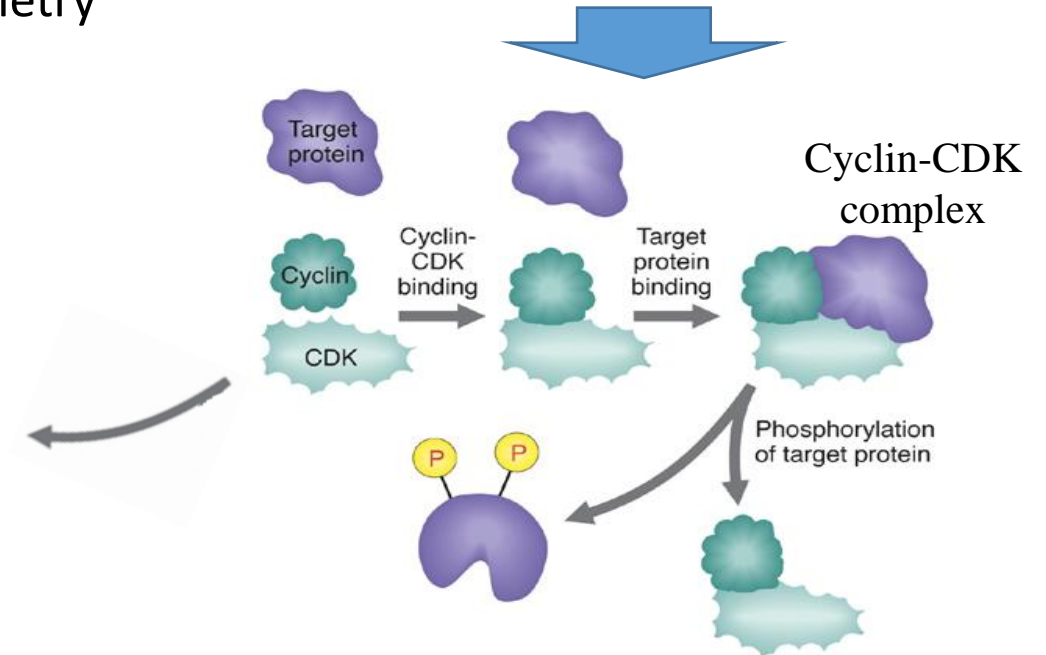
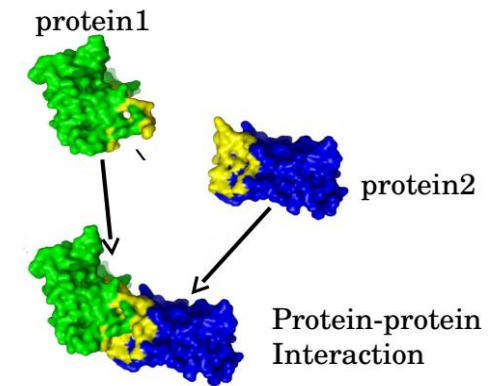
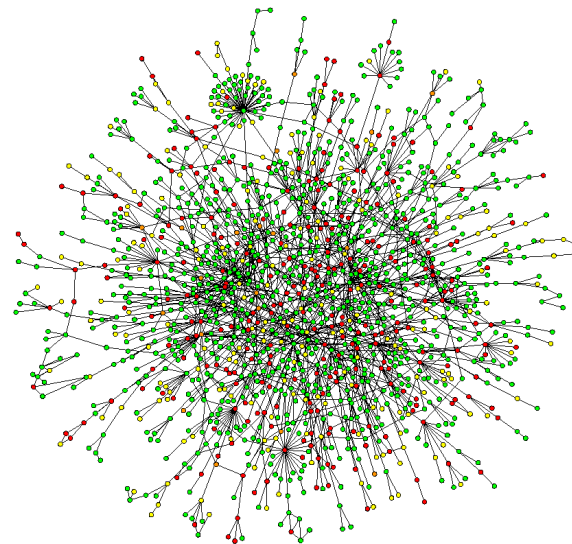
Systems Biology Resources

- Biological Networks:
 - Elements (node or vertex) + Relationships (Edges)
 - Modeled by Mathematical Graphs
- Gene regulatory Networks
- Metabolic Networks
- Protein Interaction Networks
- Co-expression Networks



Protein-Protein Interaction (PPI) Networks

- Physical Interaction between Proteins
 - Physical contacts of two or more proteins
- Detection Methods:
 - Yeast-two hybrid
 - Affinity purification with mass spectrometry
 - Text mining



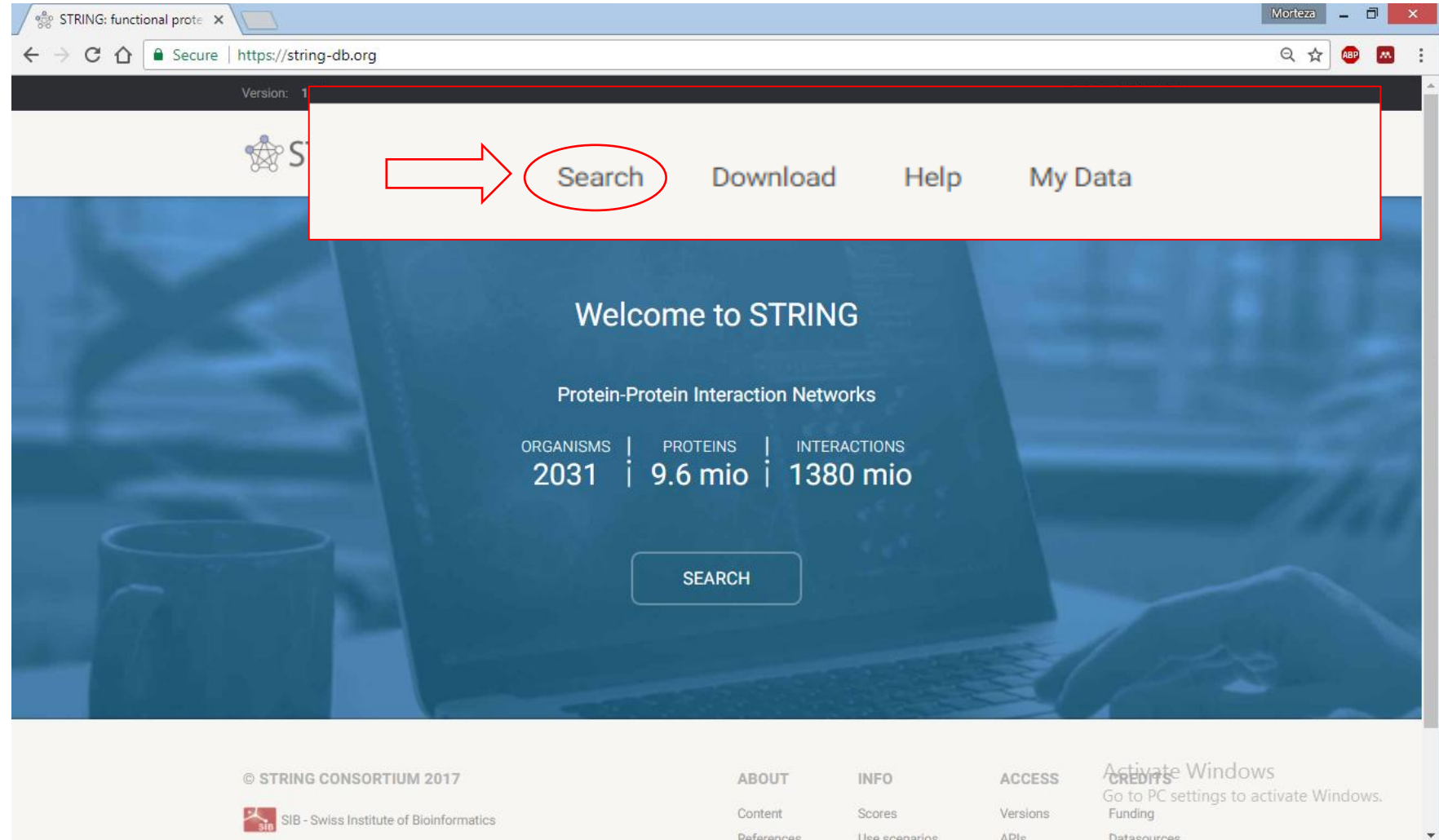
PPI Networks

- Databases:

Database name	URL	Latest update
DIP [26]	http://dip.doe-mbi.ucla.edu/dip/	2016/07/31
BioGRID [27]	https://thebiogrid.org/	2016/12/25
STRING [28]	http://string-db.org/	2016/04/16
IntAct [29]	http://www.ebi.ac.uk/intact/	2016/12/13
HIPPIE [30]	http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/index.php	2016/06/24



String Database



String Database

STRING: functional prote...

Secure | https://string-db.org/cgi/input.pl?UserId=QG6Q2Z1NEcgR&sessionId=rm6D6ITbV136&input_page_show_search=on

Version: 10.5 LOGIN REGISTER

SEARCH

Protein by name >

Protein by sequence >

Multiple proteins >

Multiple sequences >

Organisms >

Protein families ("COGs") >

Examples >

Random entry >

Single Protein by Name / Identifier

Protein Name: (examples: #1 #2 #3)

Organism: auto-detect ▼

SEARCH

© STRING CONSORTIUM 2017

SIB - Swiss Institute of Bioinformatics

ABOUT INFO ACCESS Credits

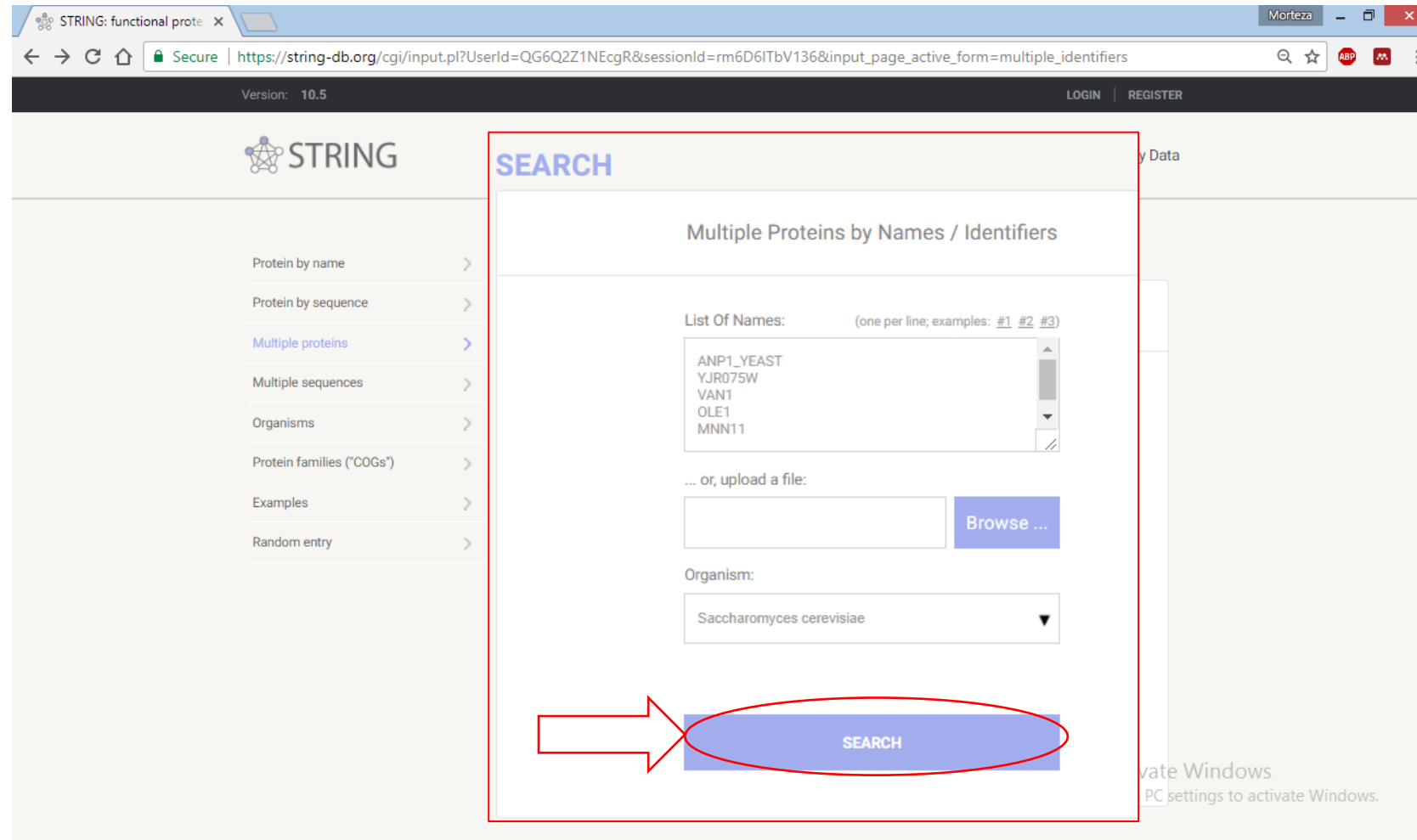
Content Scores Versions

References Use scenarios APIs Datasources

Activate Windows
Go to PC settings to activate Windows.



String Database



STRING: functional prote

Secure | https://string-db.org/cgi/input.pl?UserId=QG6Q2Z1NEcgR&sessionId=rm6D6lTbV136&input_page_active_form=multiple_identifiers

Version: 10.5 LOGIN REGISTER

STRING

Protein by name >

Protein by sequence >

Multiple proteins >

Multiple sequences >

Organisms >

Protein families ("COGs") >

Examples >

Random entry >

SEARCH

Multiple Proteins by Names / Identifiers

List Of Names: (one per line; examples: #1 #2 #3)

ANP1_YEAST
YJR075W
VAN1
OLE1
MNN11

... or, upload a file:

Browse ...

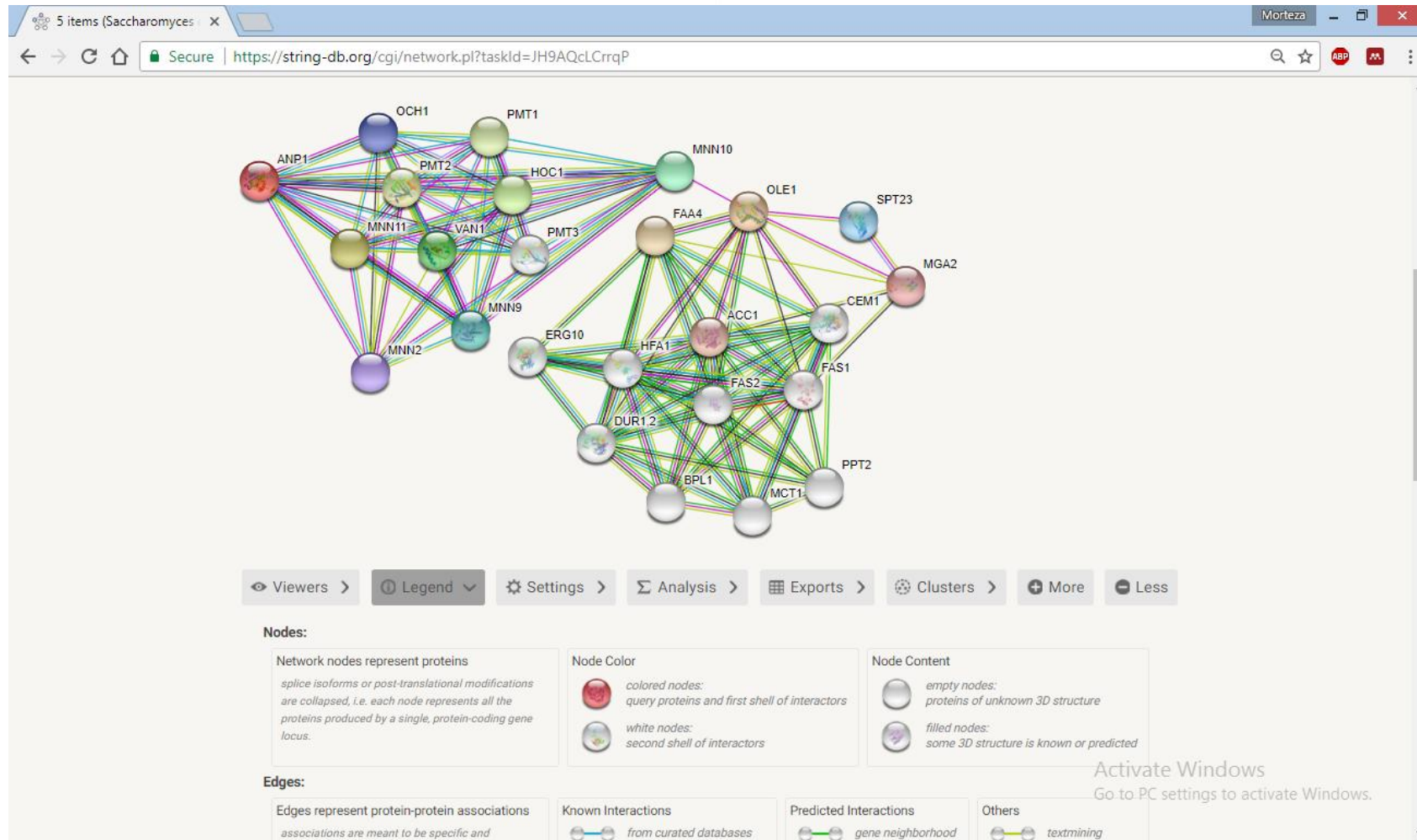
Organism:

Saccharomyces cerevisiae

SEARCH



String Database



BioGRID Database

The screenshot shows the BioGRID 3.4 website. The header includes the BioGRID logo and navigation links: home, help, wiki, tools, contribute, stats, downloads, partners, and about us. The main content area features a welcome message and a search bar. The search bar has a dropdown menu set to 'Homo sapiens' and a 'SUBMIT GENE SEARCH Q' button. Below the search bar are links for 'Advanced Search', 'Search Tips', and 'Featured Datasets'. The left sidebar contains links for 'INTERACTION STATISTICS' and 'LATEST DOWNLOADS'. The bottom section is divided into two columns: 'AREAS OF INTEREST TO HELP YOU GET STARTED' and 'BIOGRID FUNDING AND PARTNERS'. The first column includes links for 'Build and Download Interaction Datasets', 'Link To Us or Submit Interactions', 'Online Tools and Resources', and 'View Our Interaction Statistics'. The second column displays logos for various funding and partner organizations, including NIH, CIHR IRSC, Genome Québec, Mount Sinai Hospital, Princeton University, Université de Montréal, SGD, University of Edinburgh, and IMEx. A 'more partners' link is also present.

BioGRID 3.4

home help wiki tools contribute stats downloads partners about us

Welcome to the Biological General Repository for Interaction Datasets

BioGRID is an interaction repository with data compiled through comprehensive curation efforts. Our current index is version 3.4.155 and searches 63,959 publications for 1,507,991 protein and genetic interactions, 27,785 chemical associations and 38,559 post translational modifications from major model organism species. All data are **freely** provided via our search index and available for download in standardized formats.

INTERACTION STATISTICS LATEST DOWNLOADS

Search the BioGRID

Search by identifiers, keywords, and gene names...

Homo sapiens

SUBMIT GENE SEARCH Q

Advanced Search Search Tips Featured Datasets

By Gene By Publication

AREAS OF INTEREST TO HELP YOU GET STARTED

Build and Download Interaction Datasets

Create custom interaction datasets by protein or by publication. You can also download our entire dataset in a wide variety of standard formats.

Link To Us or Submit Interactions

Send us your datasets or link to the BioGRID directly from your own website or database. Full details on how to contribute are available here.

Online Tools and Resources

We've developed tools that make use of BioGRID data. Check out the list of tools to see if we can help you work with our data.

View Our Interaction Statistics

Find out how many organisms, proteins, publications, and interactions are available in the current release of the BioGRID.

BIOGRID FUNDING AND PARTNERS

NIH CIHR IRSC Genome Québec

MOUNT SINAI HOSPITAL PRINCETON UNIVERSITY Université de Montréal

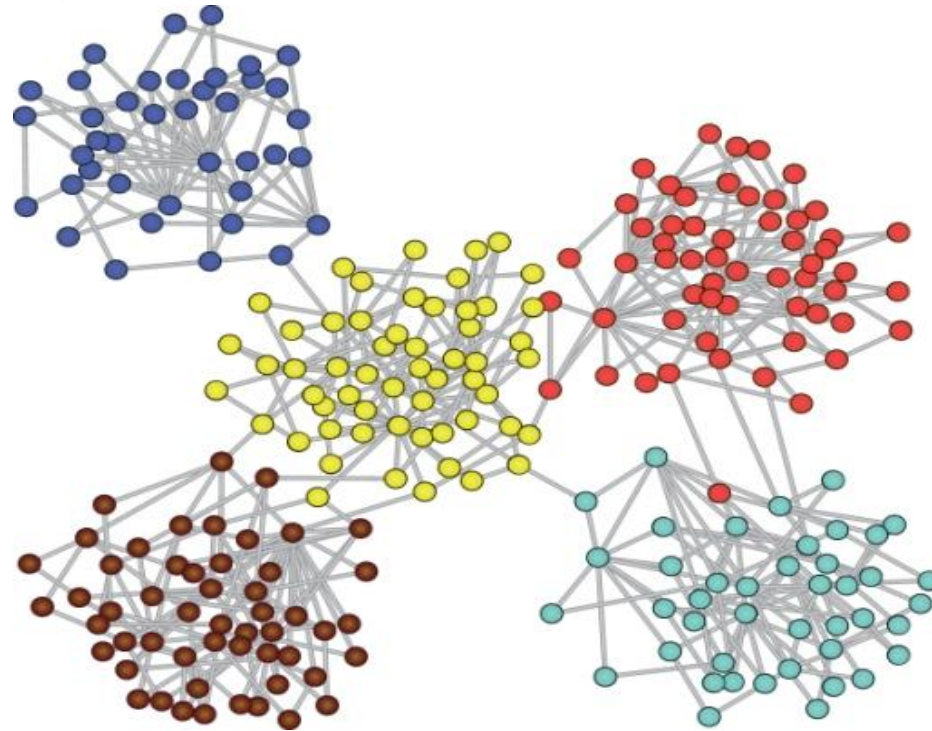
SGD University of Edinburgh IMEx

more partners

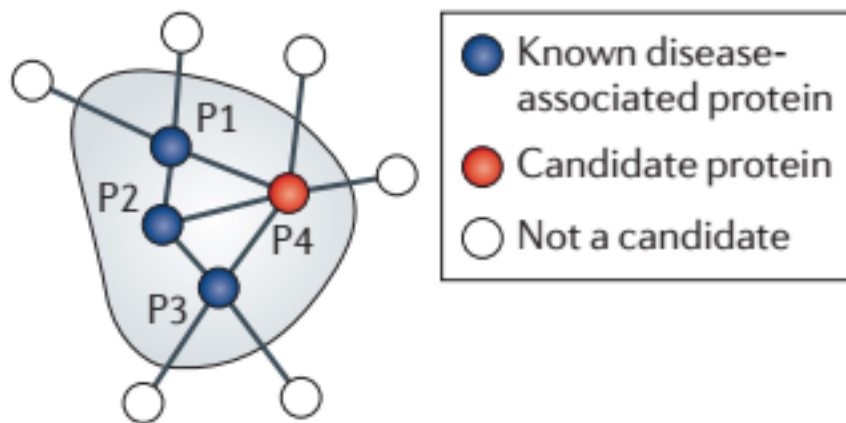
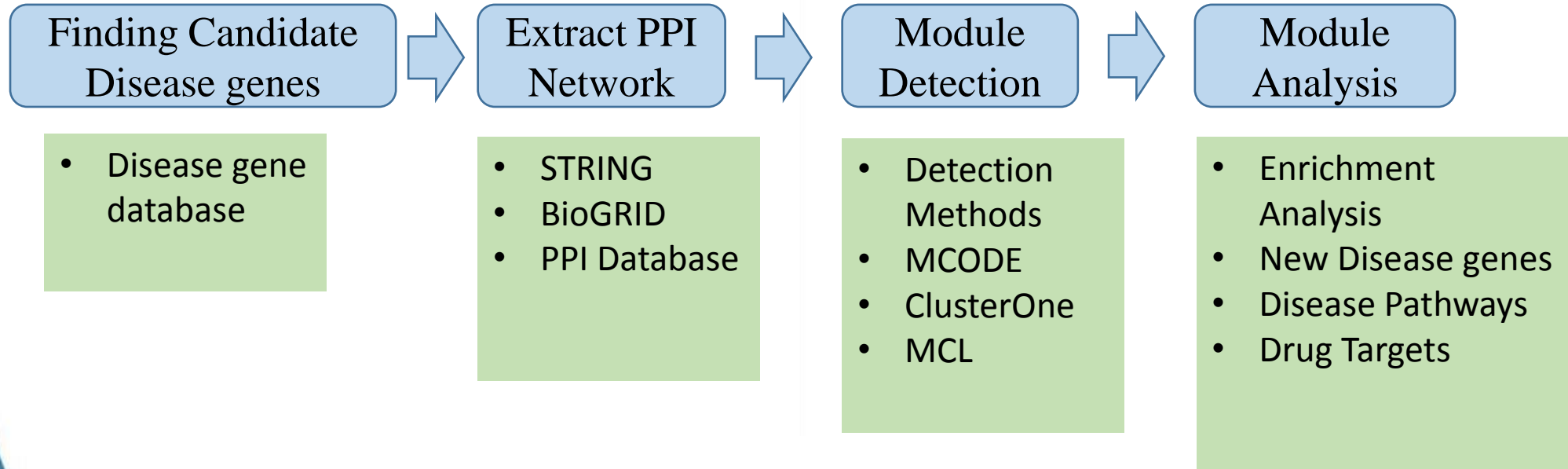


PPI Networks Analysis

- Modules or Clusters in PPI:
 - Basic definition: **Dens subnetworks (highly connected)**
 - Biological Definition: **A group of proteins that shared a biological process (Protein Complexes)**



Modules Analysis Workflow



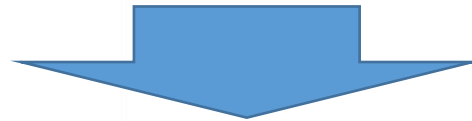
Module Detection Methods

- Using Topological Properties of the Network:
 - Finding **Cliques** (Fully Connected Subnetworks): **Cfinder**
 - Finding **Dense Subnetworks**: **MCODE**, **ClusterOne**
 - Simulating **Random Walks** on the Network: **MCL**
- Integrating Biological Information:
 - Integrate **Gene Ontology** Information: **WCOACH**
 - Finding **Active Modules** by Integrating Expression Data: **JactiveModules**

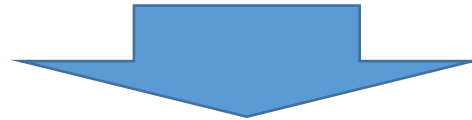


Gene Set Enrichment Analysis

Is a **set of genes** (or gene products) **significantly** involved in a **biological process or pathway**?



Search these **genes** in a **biological process or pathway database** and calculate a **P-Value**



Gene Set Enrichment



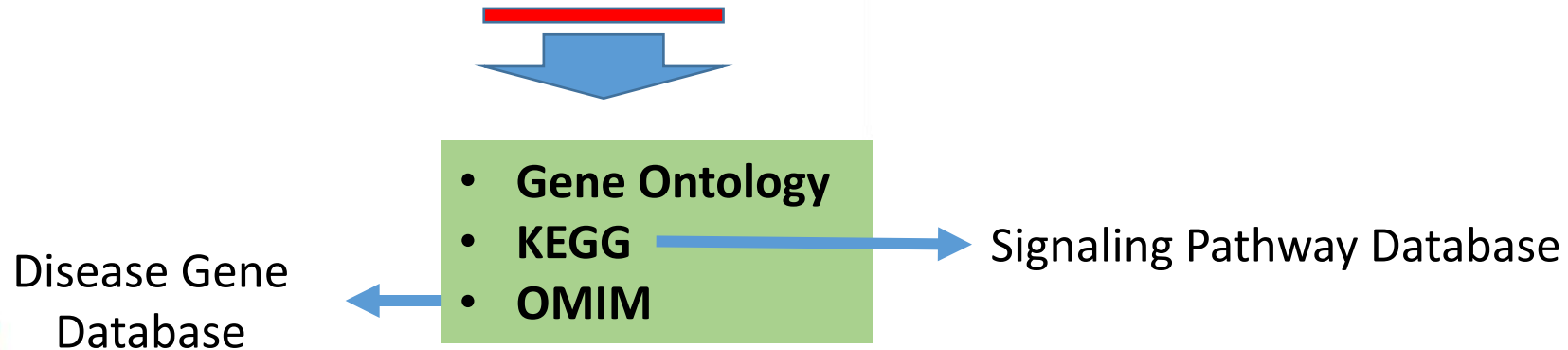
Gene Set Enrichment Analysis

- Input:

A set of genes or proteins (A Module)

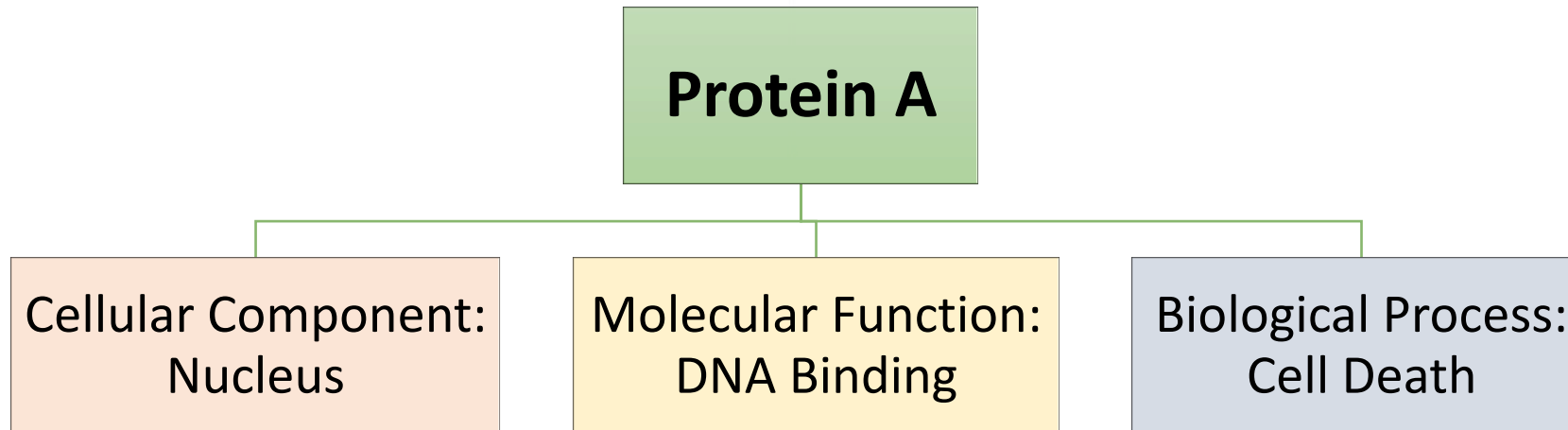
- Output:

Common Biological Processes or Pathways or Diseases based on a **Database** with P-Value



Gene Ontology Database

- Answer to three question about a Protein (or Gene):
 - What is the **location** of a protein in the cell? **Cellular Component (CC)** of a protein
 - What is the **Function** of a protein? **Molecular Function (MF)** of a Protein
 - What **Process** does a Protein involve? **Biological Process (BP)** of a Protein



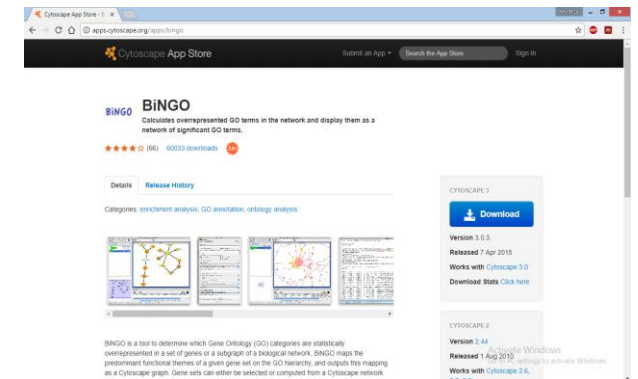
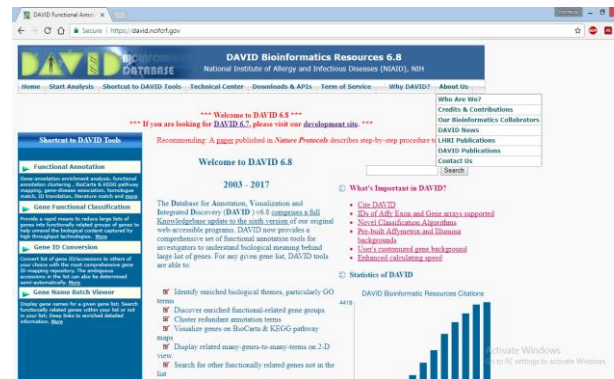
Enrichment Analysis Tools

- Online Tools

- DAVID: <https://david.ncifcrf.gov/>
- GeneCodis: <http://genecodis.cnb.csic.es/analysis>

- Offline Tools:

- BiNGO: <http://apps.cytoscape.org/apps/bingo>



Enrichment Analysis Tools

Analysis | Comparative Analysis | Web Services | Help | Release Info

GeneCodis
Gene annotations
co-occurrence discovery

1. Select the organism [?]
2. Select the annotations [?] [Last update on Dec, 2011]

Annotations

☒ GO Biological Process
☒ GO Molecular Function
3. Paste your lists of genes [see allowed IDs]

☒ Paste list of genes [?]

A2M
ABL1
APBA1
APBB1
APLP1
APLP2

☐ Or upload a file with the list of genes [?]

Choose File No file chosen
- + 5. Use your own annotations

Email (optional):

Job title (optional):

Interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Research* 2009, doi: 10.1093/nar/gkp416 [PubMed]

Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A: GENE CODIS: A web-based tool for finding significant concurrent annotations in gene lists. *Genome Biology* 2007 8(1):R3 [PubMed]

Please, submit suggestions, comments and bug reports to: fbio@cnb.csic.es



Enrichment Analysis Tools

ting Sta x android ap x How to Sel x Installing B x React Nati x Google Tra x GeneCoDis x betweenne x betweenne x Morteza

genecondis.cnb.csic.es/analysis/job-2630967444783

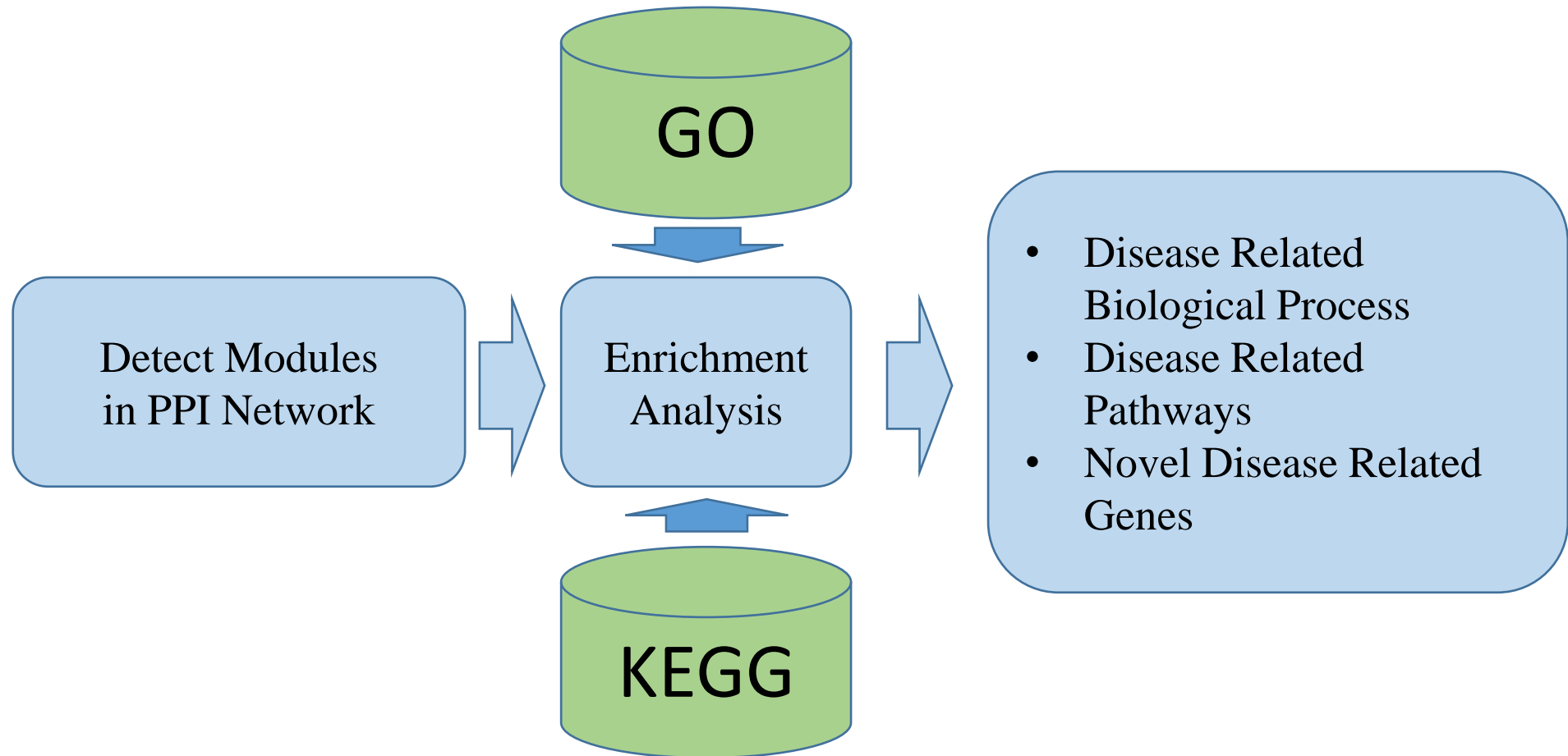
NGR = Number of annotated genes in the reference list;
 TNGR = Total number of genes in the reference list;
 NG = Number of annotated genes in the input list;
 TNG = Total number of genes in the input list;
 Hyp = Hypergeometric pValue; Hyp* = Corrected hypergeometric pValue

Filter the results by annotation name

Genes	NGR	TNGR	NG	TNG	Hyp	Hyp*	Annotations
28 genes	1072	34208	28	50	2.5258e-29	4.11706e-26	GO:0005515: protein binding (MF) GO:0005886: plasma membrane (CC)
41 genes	4463	34208	41	50	3.44292e-28	2.80598e-25	GO:0005515: protein binding (MF)
20 genes	324	34208	20	50	6.81213e-28	3.70126e-25	(KEGG) 05200: Pathways in cancer
17 genes	226	34208	17	50	3.83514e-25	1.56282e-22	GO:0005515: protein binding (MF) (KEGG) 05200: Pathways in cancer
14 genes	138	34208	14	50	1.27228e-22	4.14763e-20	GO:0005634: nucleus (CC) GO:0005515: protein binding (MF) (KEGG) 05200: Pathways in cancer



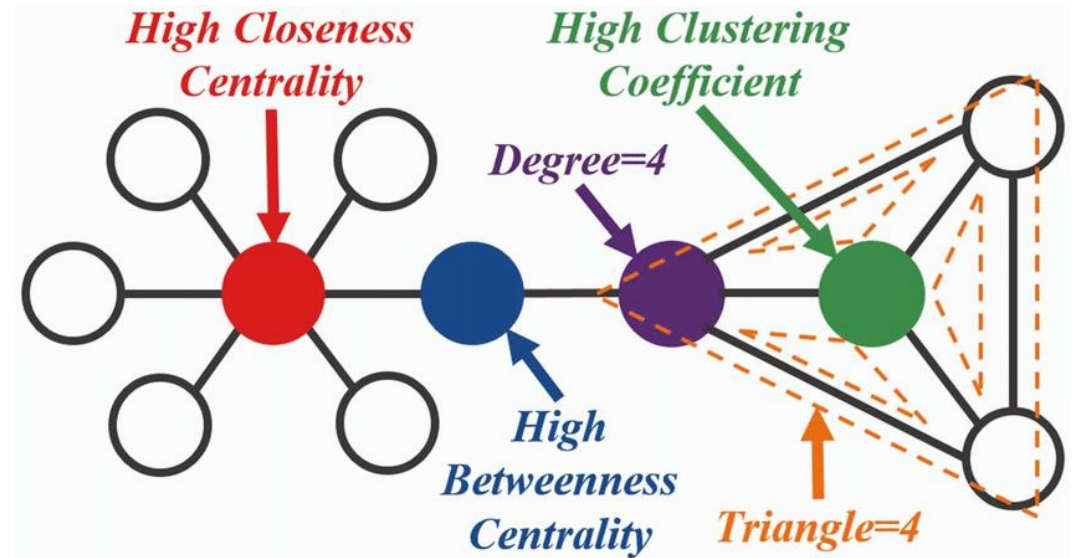
PPI Network Module Analysis



Centrality Measures and Finding Key Proteins

- Centrality Measure: A Metric to **rank nodes** based on specific topological property in a network
 - Degree Centrality: Finding Hub Proteins
 - Betweenness Centrality: Finding Between Cluster Proteins
 - Clustering Coefficient

Key Proteins can be introduced as **Drug Targets**

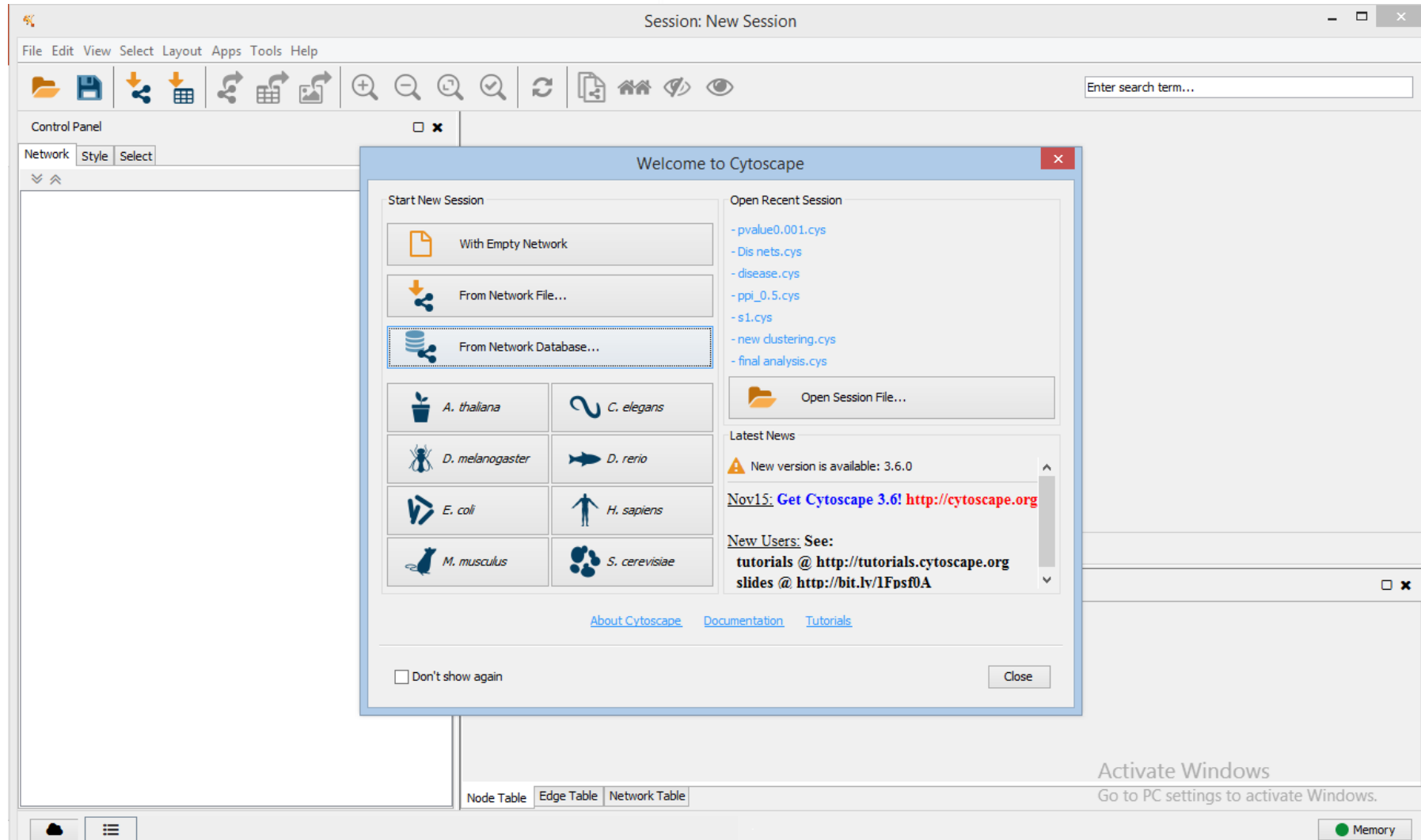


Cytoscape: A Powerful Framework for Systems Biology Analysis

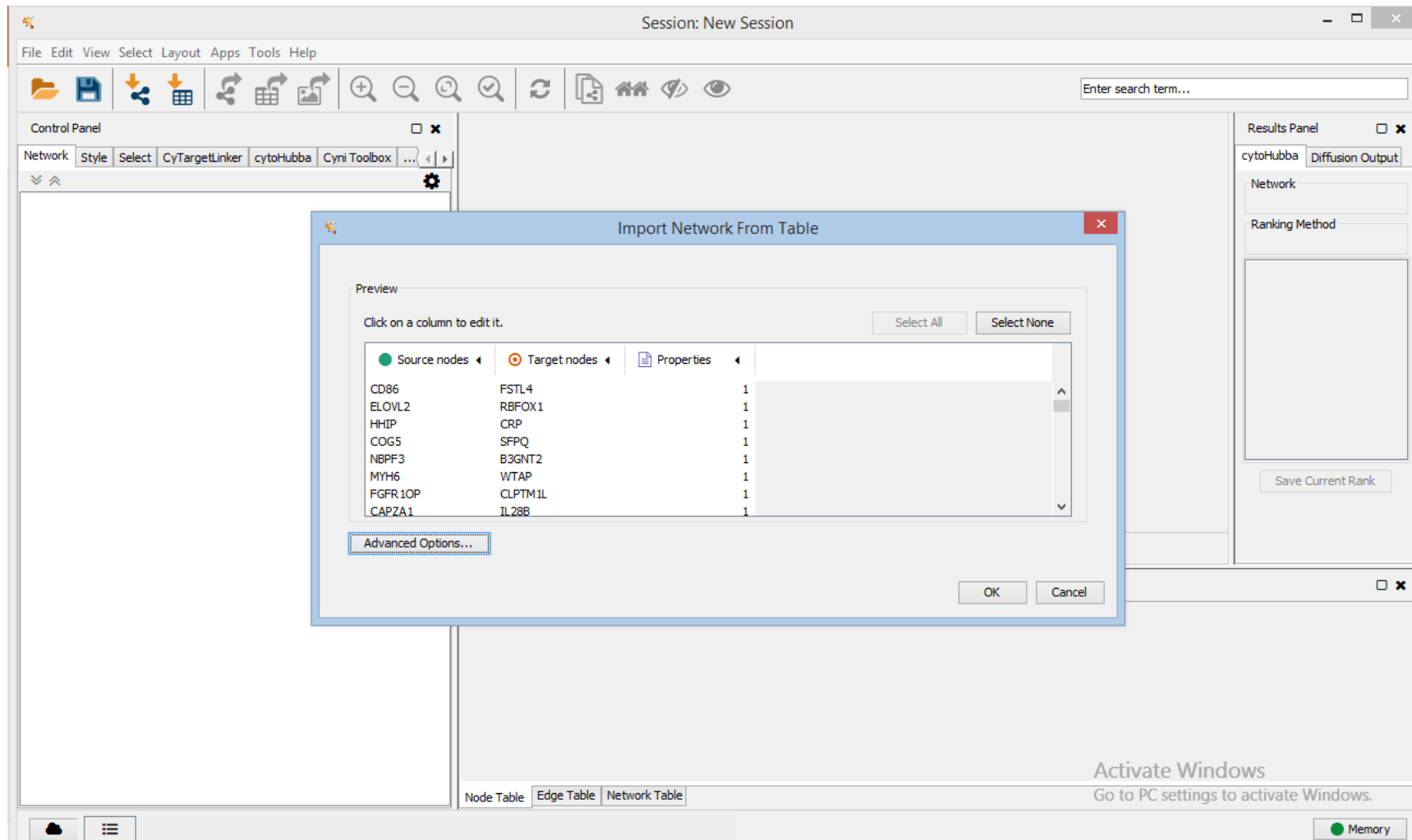
- Installing Requirement:
 - Java
- Easy Installation on any OS
- Find and add many Apps to Analysis Biological Networks



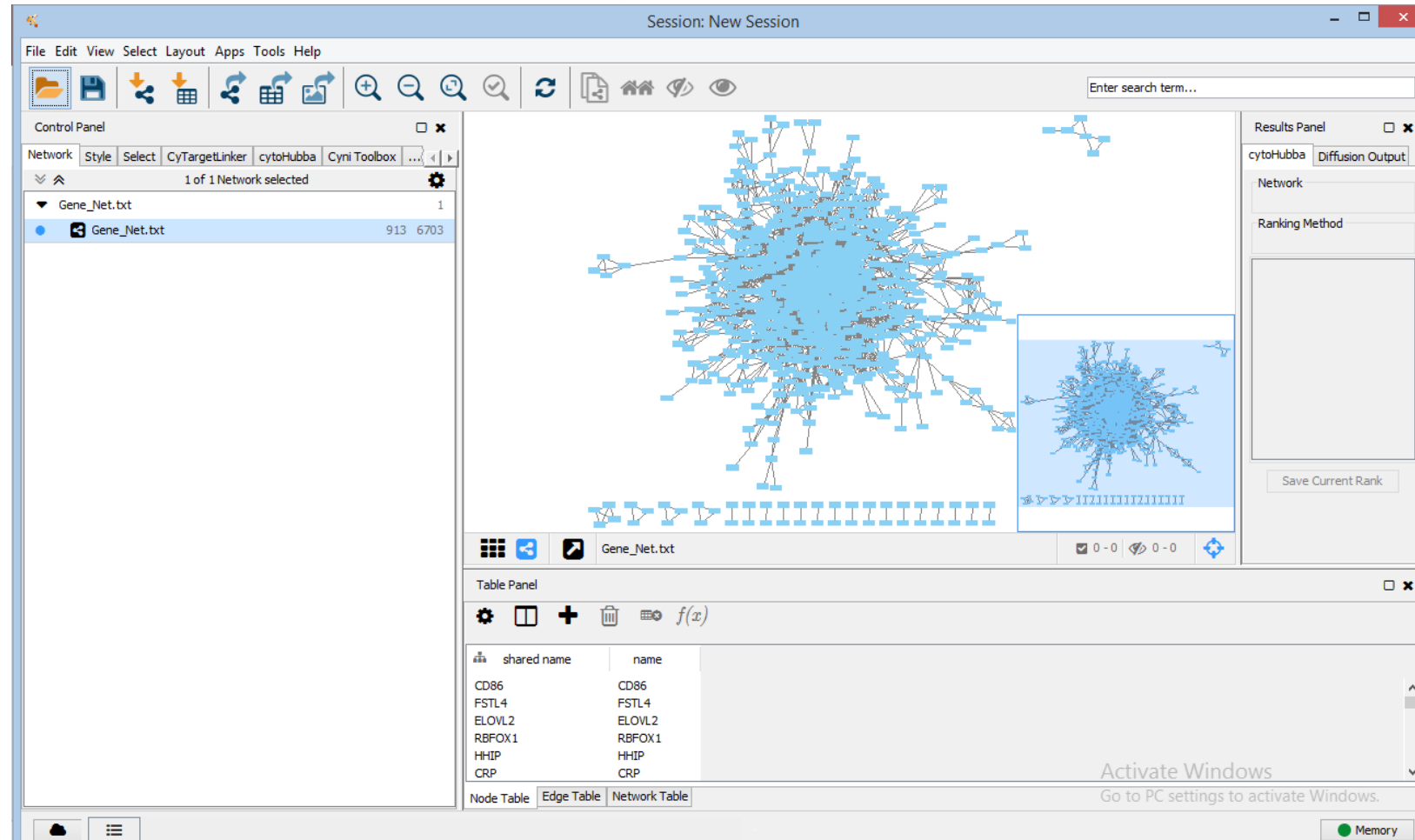
Cytoscape: Welcome Screen



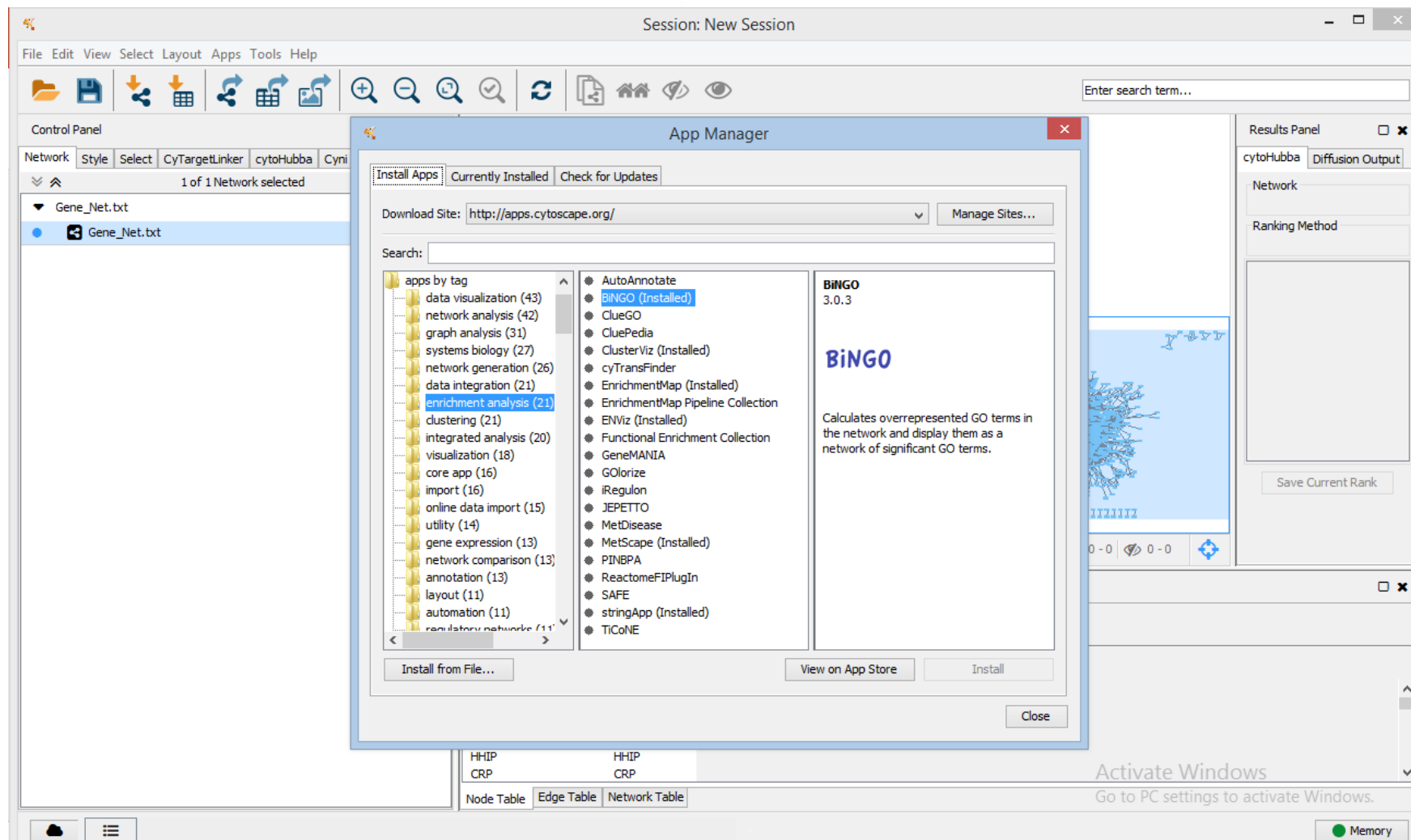
Cytoscape: Import Network from File



Cytoscape: View Network



Cytoscape: App Manager



Thank you for your attention

