

Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

Supported by



Creative Commons

This page is available in the following languages:

Afrikaans Български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto
Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
Euskara Suomi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macdonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски српски (latinica) Sotho svenska
中文 華語 (台灣) isiZulu



Attribution-Share Alike 2.5 Canada

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work



Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

[Disclaimer](#)

Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
[English](#) [French](#)

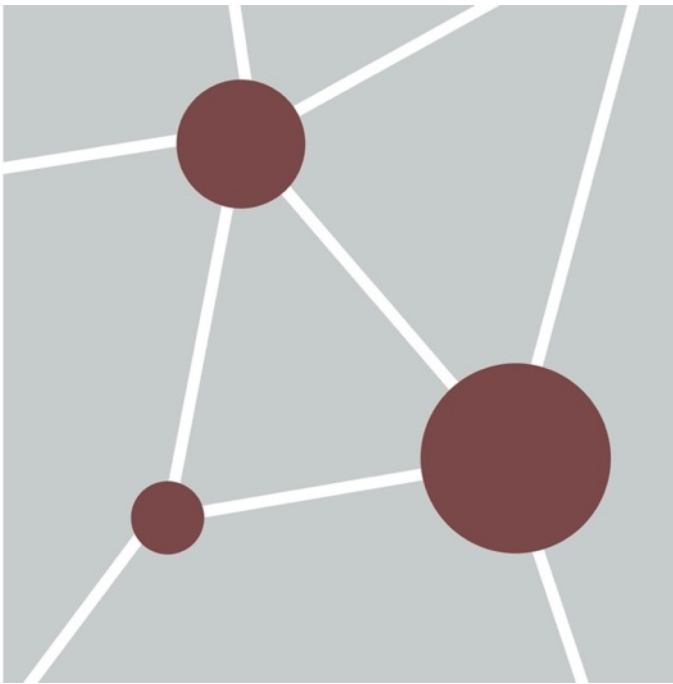
[Learn how to distribute your work using this licence](#)

In-depth Analysis of Networks and Pathways

Robin Haw

Pathway and Network Analysis of -omics Data

July 27-29, 2020



Learning Objectives

- Understand the principles of pathway and network analysis.
 - Sources of pathway and network data.
 - Analytical approaches to data analysis, visualization and integration.
 - Applications of the Reactome Functional Interaction (ReactomeFI) network.

What is Pathway/Network Analysis?

- Any analytic technique that makes use of biological or molecular network information to gain insights into a biological system.
- A rapidly evolving field.
- Many approaches.

Why Pathway/Network Analysis?

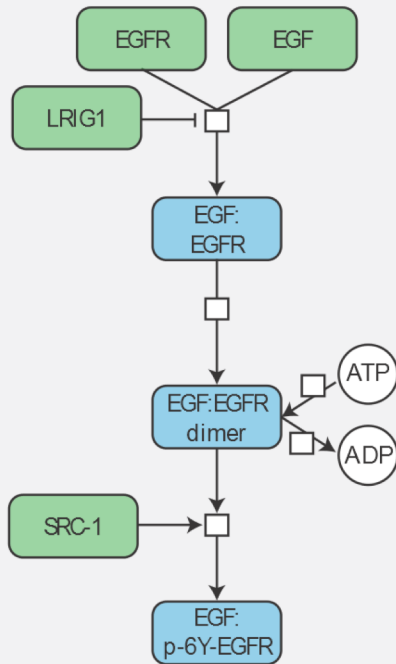
- Intuitive to Scientists:
 - Intuitive display for biological data.
 - Visualize multiple data types on a network.
 - Increase statistical power by reducing multiple hypotheses.
 - Computational methods available to automate analysis.
- Network Analysis satisfy common “use cases” in biological research:
 - Identifying hidden patterns in gene lists.
 - Creating emergent models to explain experimental observations.
 - Predicting the function of unannotated genes.
 - Establishing the framework for quantitative modeling.
 - Assisting in the development of molecular signatures.

List of Cancer Driver Genes

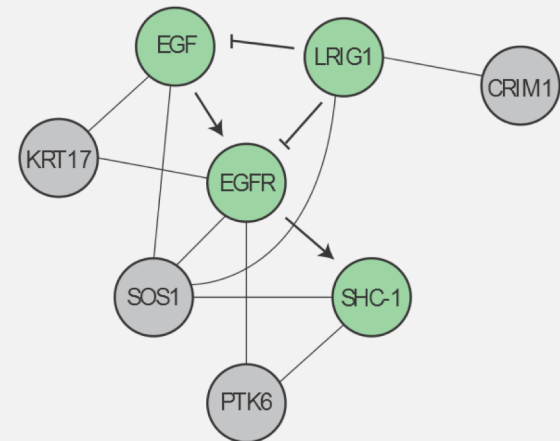
ACVR1B, ACVR2A, AJUBA, AKT1, APC, AR, ARHGAP35, ARID1A, ARID5B, ASXL1, ATM, ATR, ATRX, AXIN2, B4GALT3, BAP1, BRAF, BRCA1, BRCA2, CBF3, CCND1, CDH1, CDK12, CDKN1A, CDKN1B, CDKN2A, CDKN2C, CEBPA, CHEK2, CRIPAK, CTCF, CTNNB1, DNMT3A, EGFR, EGR3, EIF4A2, ELF3, EP300, EPHA3, EPHB6, EPPK1, ERBB4, ERCC2, EZH2, FBXW7, FGFR2, FGFR3, FLT3, FOXA1, FOXA2, GATA3, H3F3C, HGF, HIST1H1C, HIST1H2BD, IDH1, IDH2, KDM5C, KDM6A, KEAP1, KIT, KRAS, LIFR, LRRK2, MALAT1, MAP2K4, MAP3K1, MAPK8IP1, MECOM, MIR142, MLL2, MLL3, MLL4, MTOR, NAV3, NCOR1, NF1, NFE2L2, NFE2L3, NOTCH1, NPM1, NRAS, NSD1, PBRM1, PCBP1, PDGFRA, PHF6, PIK3CA, PIK3CG, PIK3R1, POLQ, PPP2R1A, PRX, PTEN, PTPN11, RAD21, RB1, RPL22, RPL5, RUNX1, SETBP1, SETD2, SF3B1, SIN3A, SMAD2, SMAD4, SMC1A, SMC3, SOX17, SOX9, SPOP, STAG2, STK11, TAF1, TBL1XR1, TBX3, TET2, TGFBR2, TLR4, TP53, TSHZ2, TSHZ3, U2AF1, USP9X, VEZF1, VHL, WT1

Pathways vs Networks

EGFR-centered
Pathway



EGFR-centered
Network

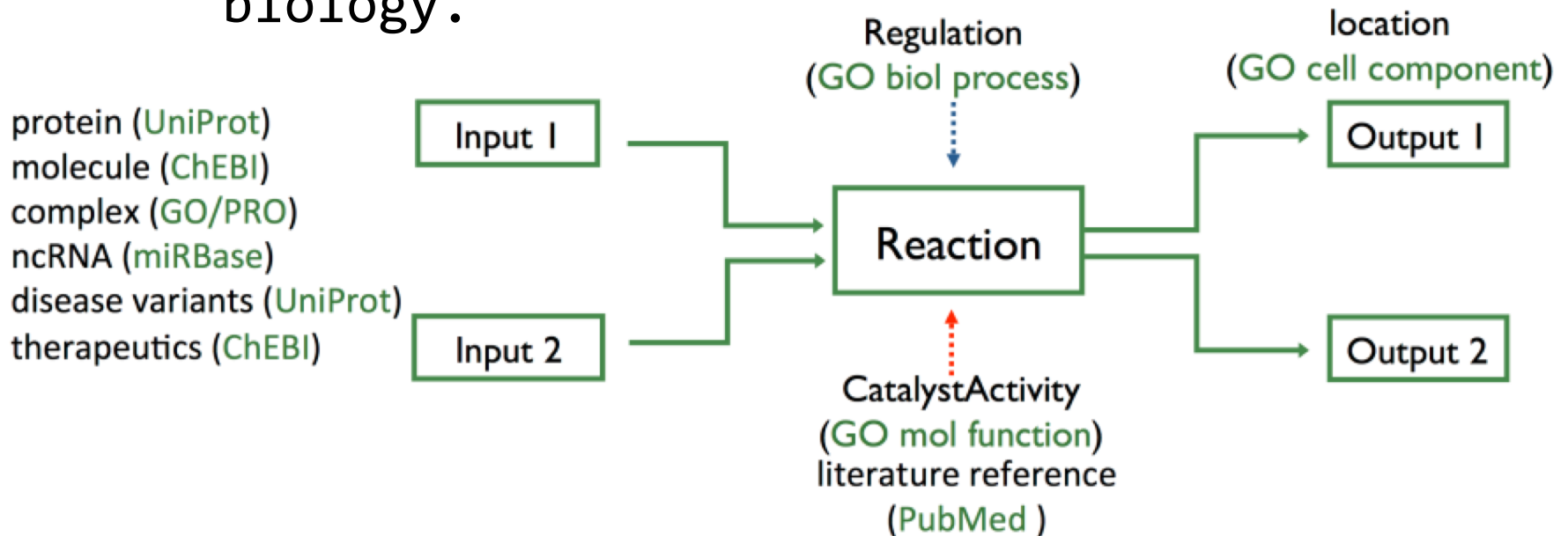


Pathway Databases

- Advantages:
 - Usually curated.
 - Biochemical view of biological processes.
 - Cause and effect captured.
 - Human-interpretable visualizations.
- Disadvantages:
 - Sparse coverage of genome.
 - Different databases disagree on boundaries of pathways.

Reaction-based Pathway Databases

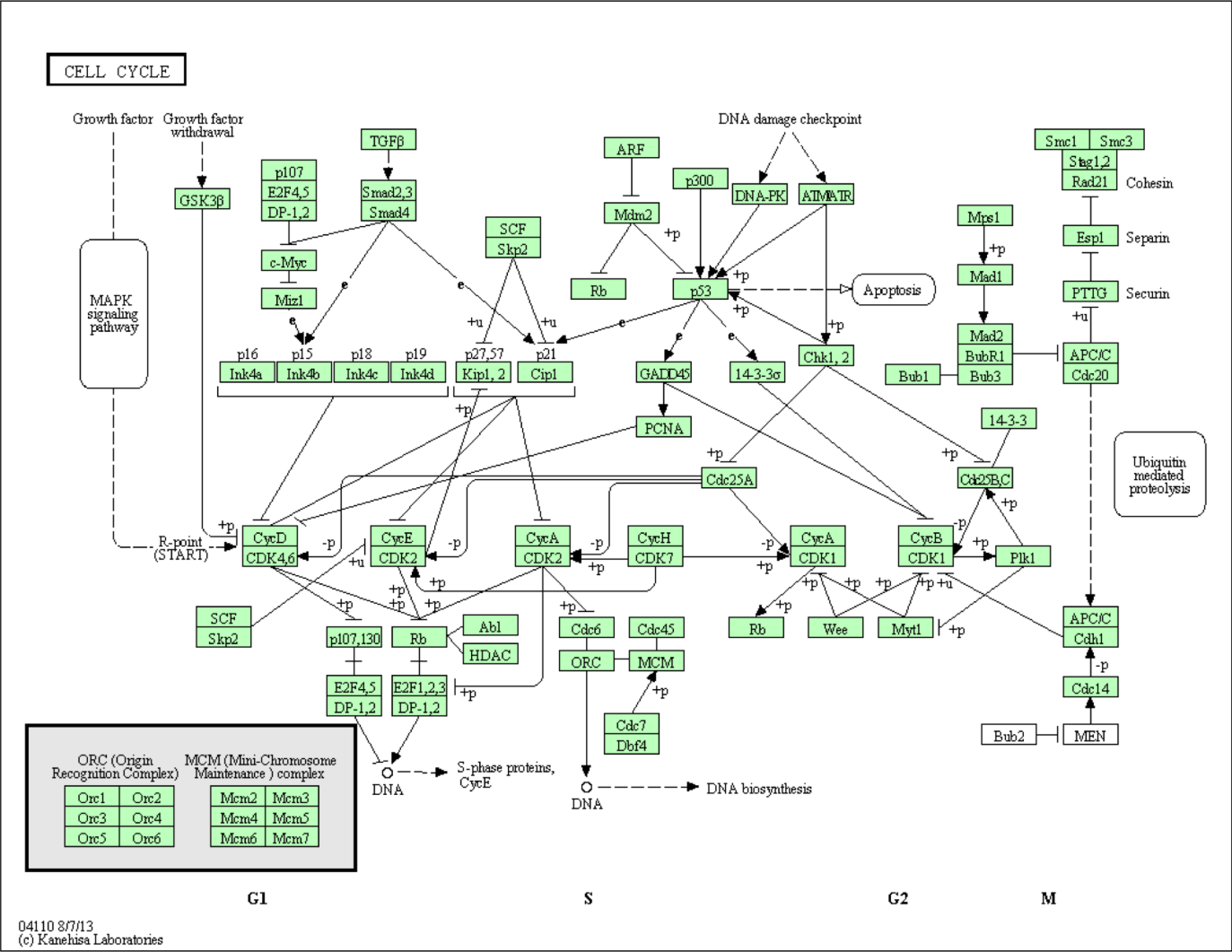
- Reactome & KEGG
 - explicitly describe biological processes as a series of biochemical reactions.
 - represents many events and states found in biology.



KEGG

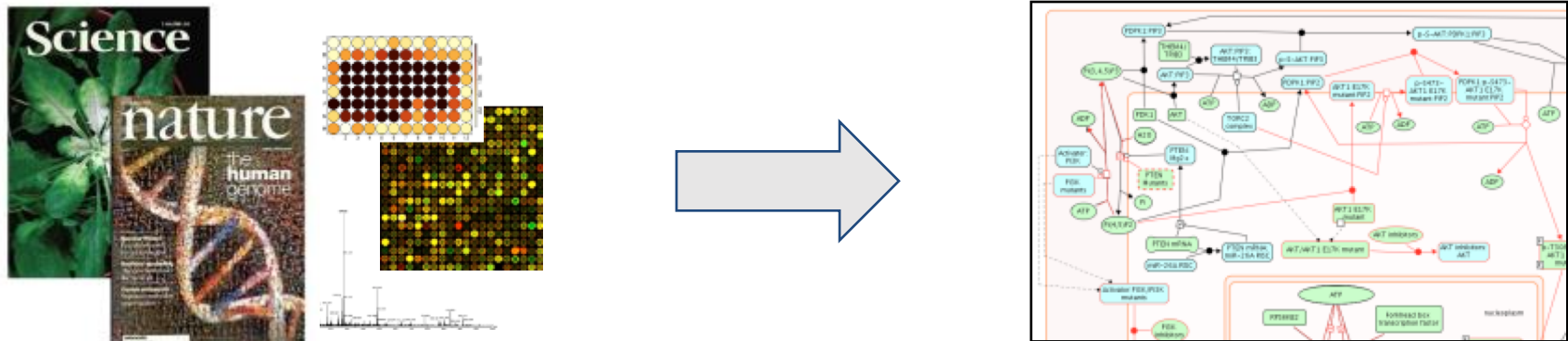
- KEGG is a collection of biological information compiled from published material → **curated database**.
- Includes information on genes, proteins, metabolic pathways, molecular interactions, and biochemical reactions associated with specific organisms
- Provides a relationship (map) for how these components are organized in a cellular structure or reaction pathway.

KEGG Pathway Diagram



Reactome

- Open source and open access pathway database
- Curated human pathways encompassing metabolism, signaling, and other biological processes.
- Every pathway is traceable to primary literature.
- Cross-reference to many other bioinformatics databases.
- Provides data analysis and visualization tools



Pathway Browser: Pathway Enrichment Analysis

reactome 3.5
Pathways for: **Homo sapiens**
Analysis:
Tour:
Layout:

Event Hierarchy:

- Developmental Biology (37/1,047) FDR: 7.1E-13
- Digestion and absorption (1/27) FDR: 3.89E-1
- Disease (62/1,147) FDR: 3.02E-13**
- DNA Repair (11/295) FDR: 4.78E-2
- DNA Replication (1/108) FDR: 8.62E-1
- Extracellular matrix organization (17/301) FDR: 1.9E-6
- Gene expression (Transcription) (42/1,42) FDR: 1.9E-6
- Hemostasis (36/743) FDR: 1.9E-6
- Immune System (60/2,229) FDR: 5.19E-1
- Metabolism (26/2,110) FDR: 9.92E-1
- Metabolism of proteins (31/2,010) FDR: 9.99E-1
- Metabolism of RNA (4/673) FDR: 9.99E-1
- Mitophagy
- Muscle contraction (1/205) FDR: 9.77E-1
- Neuronal System (7/370) FDR: 5.13E-1
- Organelle biogenesis and maintenance (4/2,110) FDR: 9.99E-1
- Programmed Cell Death (6/175) FDR: 1.02E-1
- Reproduction (5/114) FDR: 8.71E-2
- Signal Transduction (102/2,675) FDR: 4.44E-16**
- Signaling by Receptor Tyrosine Kinases
- Signaling by TGF-beta family members
- Signaling by GPCR (23/1,227) FDR: 4.71E-1
- Signaling by NOTCH (9/127) FDR: 4.21E-1
- Signaling by WNT (11/297) FDR: 4.78E-1
- Signaling by Hippo
- Signaling by Hedgehog (6/149) FDR: 8.71E-1
- Signaling by Leptin (3/11) FDR: 6.84E-3
- Integrin signaling (4/27) FDR: 9.68E-3
- Signaling by Nuclear Receptors (1/43) FDR: 1.9E-6
- MAPK family signaling cascades (23/295) FDR: 1.9E-6
- Intracellular signaling by second messengers (32/300) FDR: 0.027
- Signaling by Rho GTPases (10/416) FDR: 0.024
- Signaling by Non-Receptor Tyrosine Kinases (30/269) FDR: 0.024
- mTOR signalling (3/40) FDR: 7.48E-2

Description	Molecules	Structures	Expression	Analysis 924	Downloads				
Results for: UNIPROT (924) Type: Overrepresentation [Data: GBM Uniprot]									
Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
Signal Transduction	102	2,675	0.242	4.44E-16	1.14E-13	816	1,947	0.177	Homo sapiens
Intracellular signaling by second messengers	32	300	0.027	1.22E-15	2.5E-13	33	104	0.009	Homo sapiens
Disease	62	1,147	0.104	1.78E-15	3.02E-13	275	891	0.081	Homo sapiens
PIP3 activates AKT signaling	30	269	0.024	3.33E-15	4.86E-13	30	85	0.008	Homo sapiens
Negative regulation of the PI3K/AKT network	21	115	0.01	6.11E-15	7.82E-13	4	10	0.001	Homo sapiens
Signaling by FGFR in disease	17	72	0.007	4.52E-14	5.11E-12	74	106	0.01	Homo sapiens

127 Cancer Driver Genes

Pathway and Network Compendia

The screenshot shows the Pathway Commons website. At the top left is the Pathway Commons logo, a stylized 'P' with a green circle inside. To its right is the text 'Pathway Commons' in a large, bold, black font. Below this is the tagline 'Access and discover data integrated from public pathway and interactions databases.' in a smaller blue font. Further down, a blue link reads 'Pathway Commons, a web resource for biological pathway data.' To the right of this are navigation links for 'Data', 'Tools', 'FAQ', and 'Contact'. Below the navigation is a section titled 'Apps' which contains two search boxes. The first is labeled 'Search' and has the text 'Search the entire collection of pathways' and a search input field with the placeholder 'Names or gene IDs (e.g. 'glycolysis', 'TP53')'. The second is labeled 'PCViz' and has the text 'Get details about genes and their interactions' and a search input field with the placeholder 'Gene IDs (e.g. 'MDM2 TP53')'. Below the 'Apps' section is a section titled 'Training' which features a 'Guide' card. The 'Guide' card has a grey background and contains the text: 'A pathway analysis online textbook: Workflows provide step-by-step instruction to pathway analysis; primers for deep-dives into concepts; case studies illustrate real-world use of methods and resources.' To the right of the 'Guide' card is a network diagram with nodes and edges, overlaid on a dark background with a cityscape at night. Labels on the diagram include 'IL8- and CXCR2-mediated signaling events', 'blood coagulation', 'platelet activation', 'Thromboxan A2 receptor signaling', 'hepatocellular carcinoma', and 'immunoregulatory T cells'. The overall layout is clean and professional, with a white background and blue accents.

Pathway Commons
Access and discover data integrated from public pathway and interactions databases.

Pathway Commons, a web resource for biological pathway data.

Data Tools FAQ Contact

Apps

Search

Search the entire collection of pathways

PCViz

Get details about genes and their interactions

Training

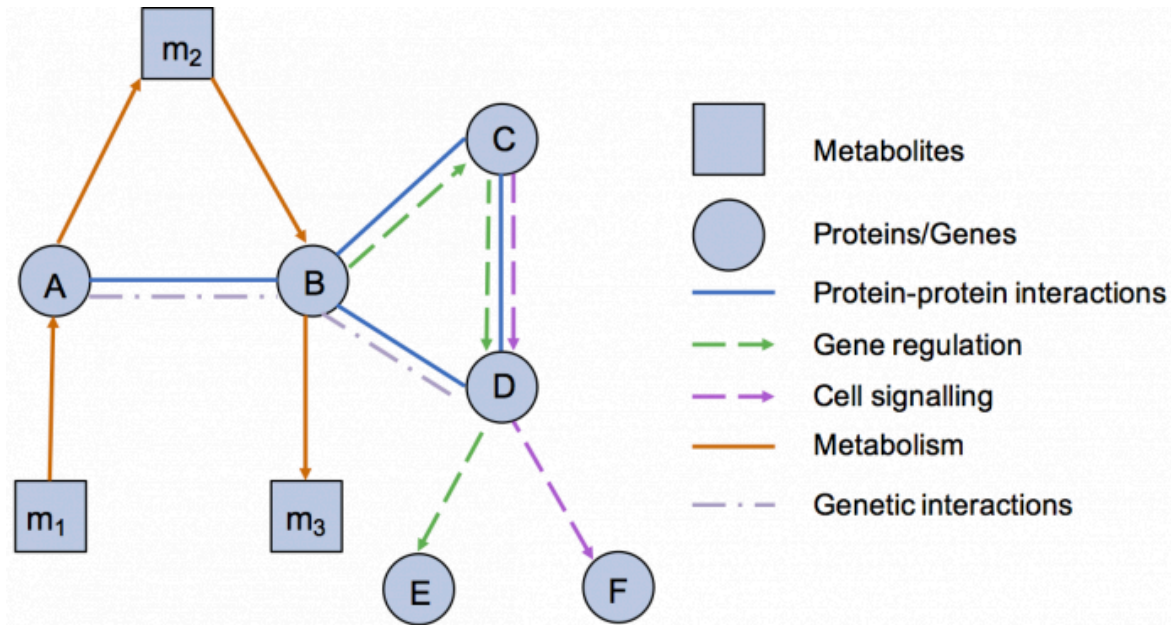
Guide

A pathway analysis online textbook: Workflows provide step-by-step instruction to pathway analysis; primers for deep-dives into concepts; case studies illustrate real-world use of methods and resources.

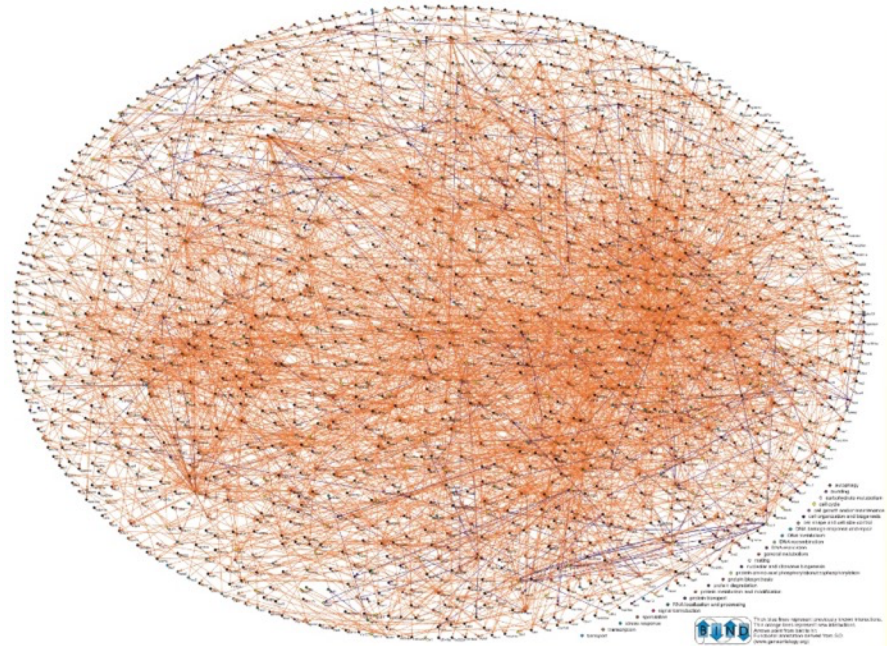
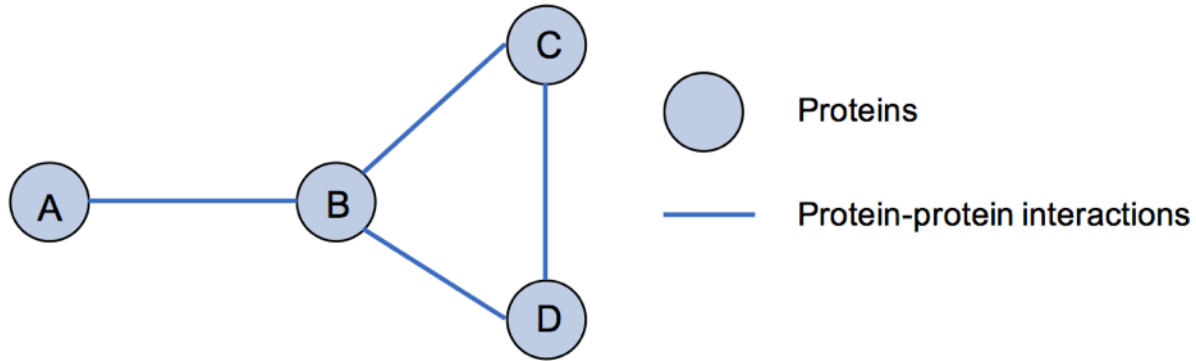
The network diagram features several nodes connected by edges. Key nodes include 'IL8- and CXCR2-mediated signaling events', 'blood coagulation', 'platelet activation', 'Thromboxan A2 receptor signaling', 'hepatocellular carcinoma', and 'immunoregulatory T cells'. The diagram is set against a dark background with a cityscape at night.

www.pathwaycommons.org

Types of Biological Networks



Protein-protein Interaction Networks



Sources of Interaction Network Data

- Databases can be built automatically or via curation.
- Popular sources of curated networks:



IMEx Partners



Data exchange formats



BioPAX

Curated Interaction Sources - Techniques

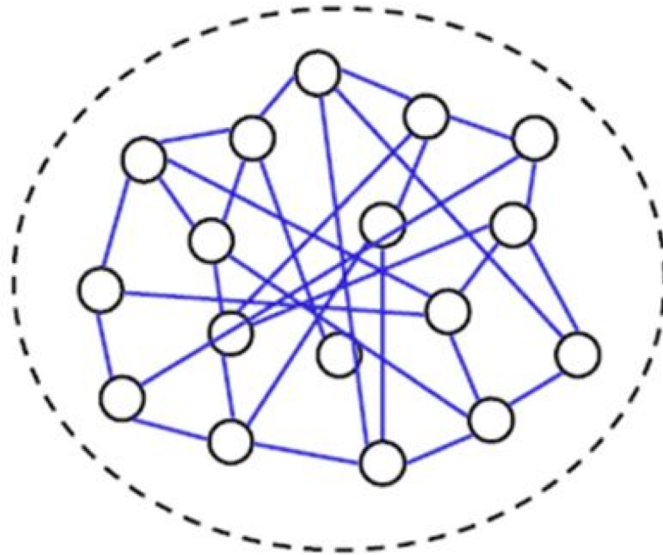
- Yeast 2 hybrid protein interactions.
- Protein complex pulldowns/mass spec.
- Genetic screens, such as synthetic lethals, enhancer/suppressor screens.
- NOT perfect:
 - Y2H interactions have taken proteins out of natural context; physical interaction = biological interaction!.
 - Protein complex pulldowns plagued by “sticky” proteins such as actin.
 - Genetic screens highly sensitive to genetic background (“network effects”).
- Computationally extract gene relationships from text, such as PubMed abstracts or PMC full text.

Network Topology



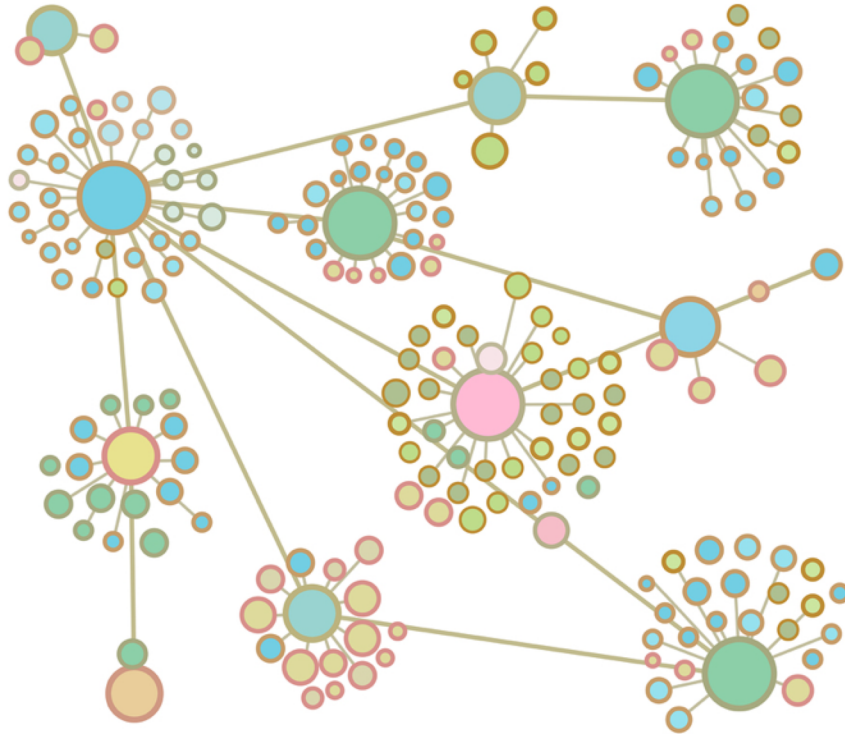
Topology reflects the way in which the nodes and edges are arranged within a network.

Small World Effect



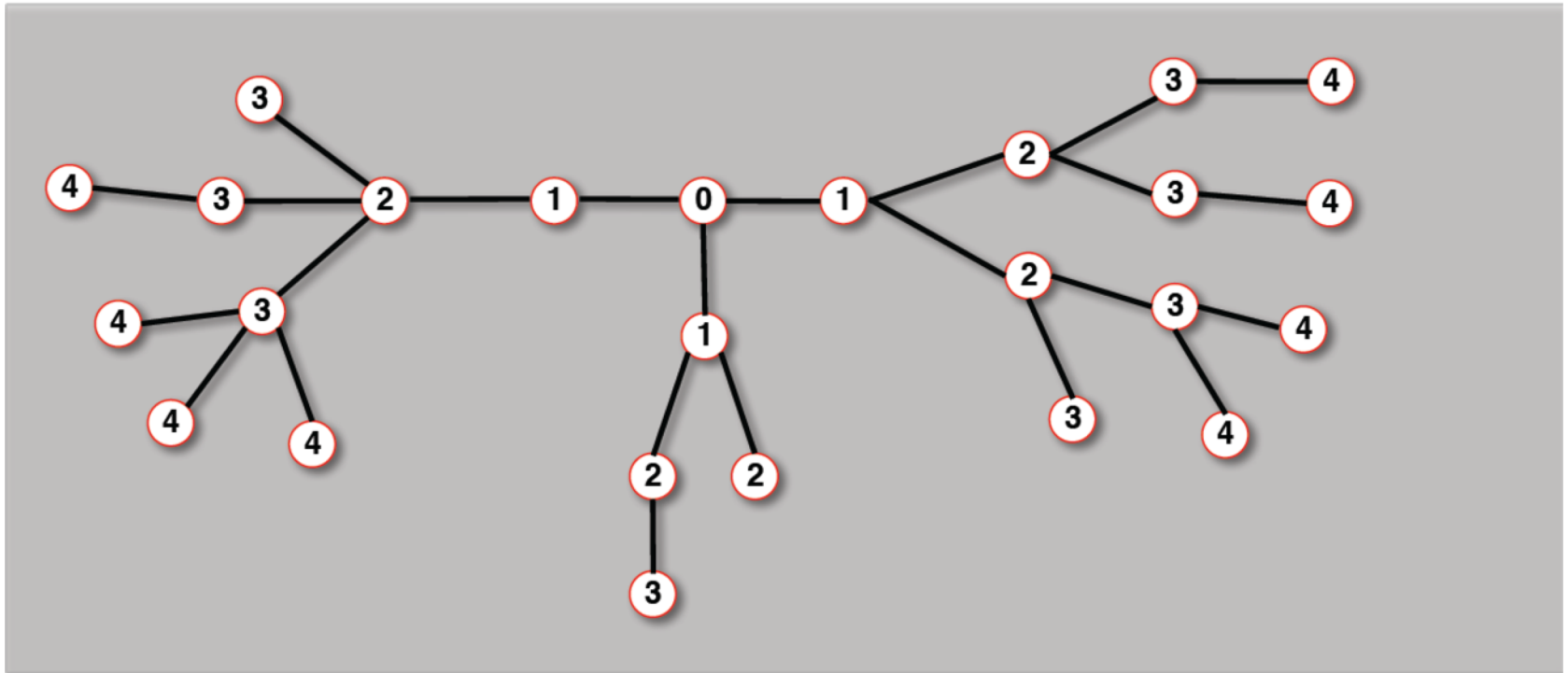
Protein-protein interaction networks show a small world effect meaning that there is great connectivity between proteins

Scale-free Networks

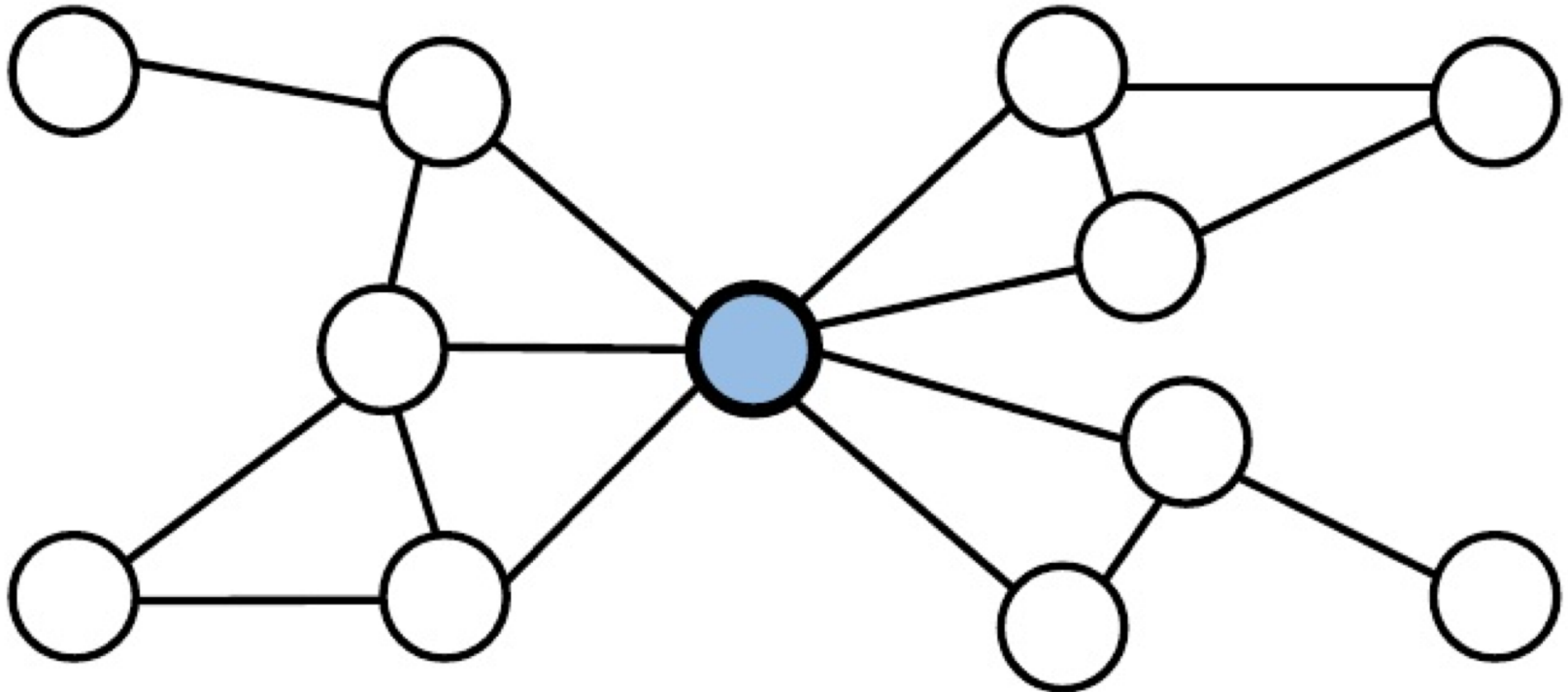


A small number of nodes with high degree and a large number of nodes with a low degree

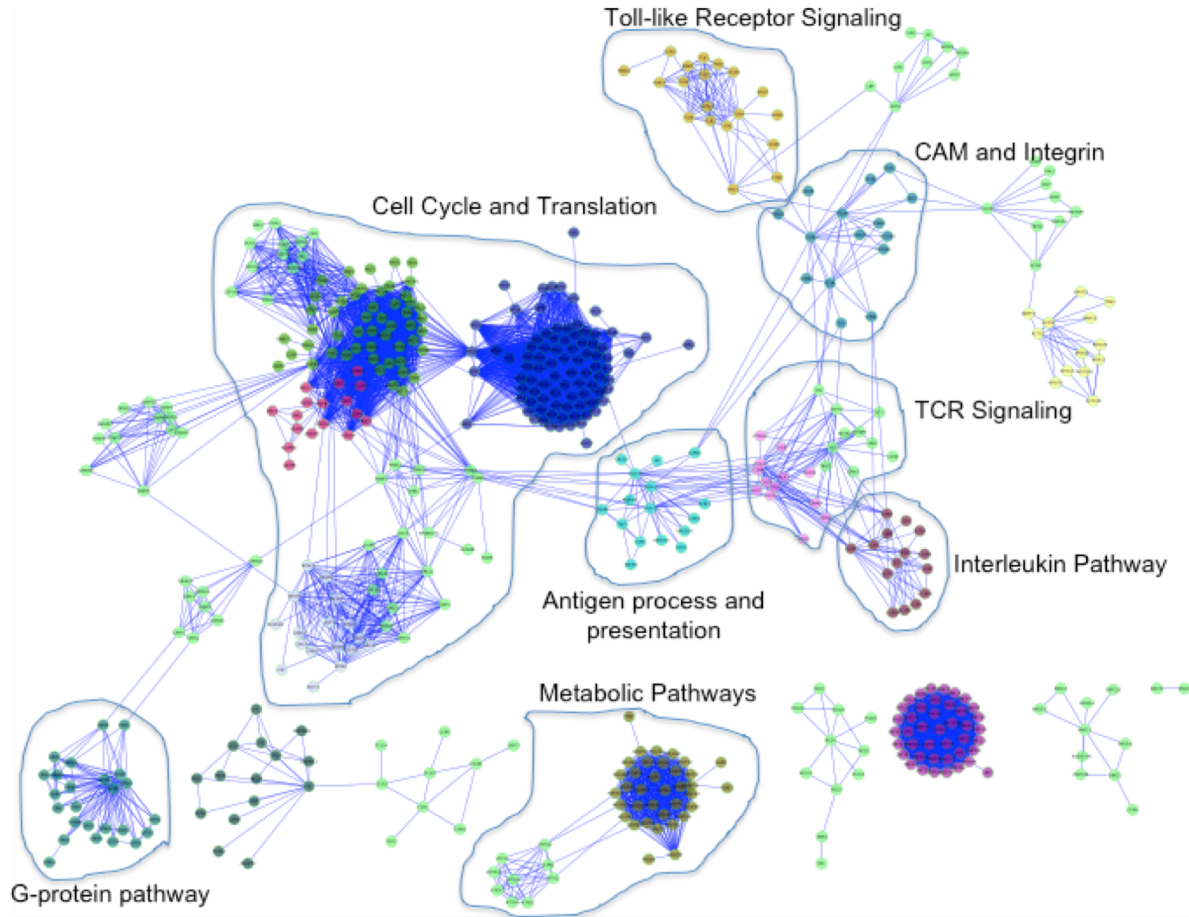
Distance in a Graph



Centrality Measures

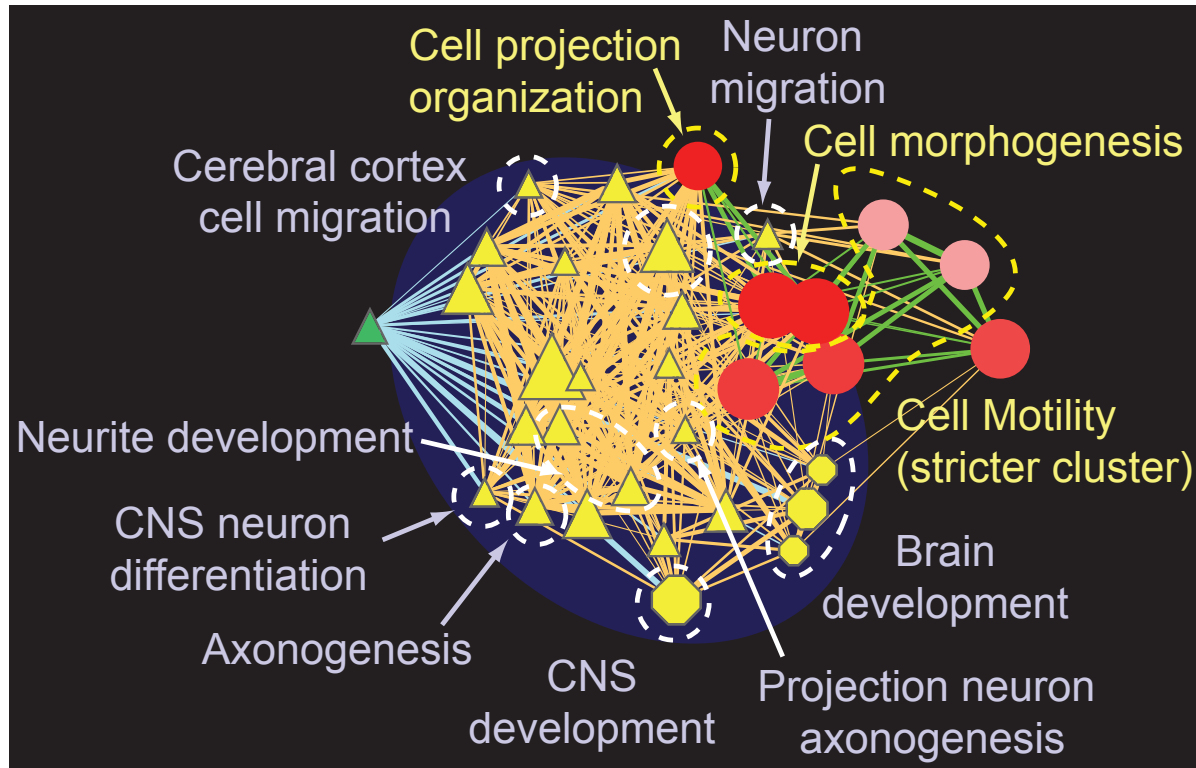


Clustering Coefficient



Transitivity or clustering coefficient of a network is a measure of the tendency of the nodes to cluster together

Annotation Enrichment Analysis



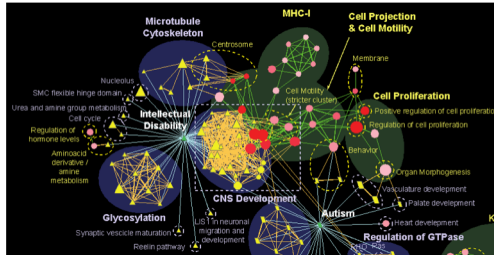
Annotation enrichment analysis using GO

Steps in Network Visualization and Analysis



1. Need a software to create the network
2. Upload our data (usually table format) to create the network
3. Navigate through the created network
4. Analyze the network (e.g do you see clusters?)
5. Label the clusters (with pathway or GO annotations)
6. Export network (table) and network image

Software to Create and Analyze Networks



Cytoscape

<http://cytoscape.org/>

Features

standalone application

open-source and free
social network / biological data

playing with visual styles

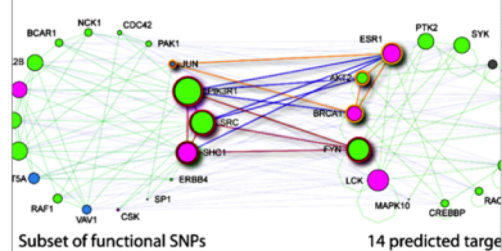
network analysis (e.g centrality)

customizable by apps



Gephi

<https://gephi.org/>

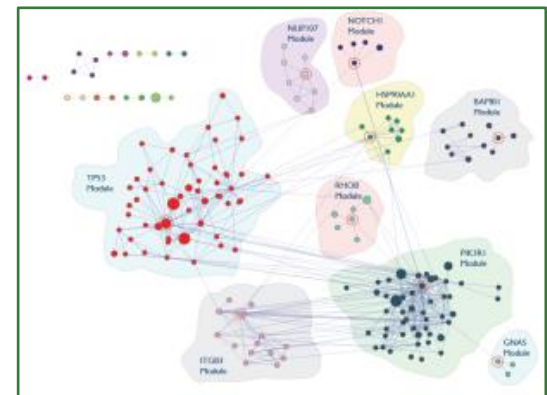
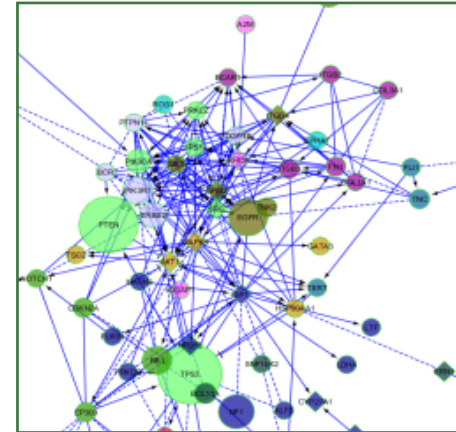


NAVIGaTOR

<http://ophid.utoronto.ca/navigator/>

Reactome Functional Interaction (FI) Network and ReactomeFIViz App

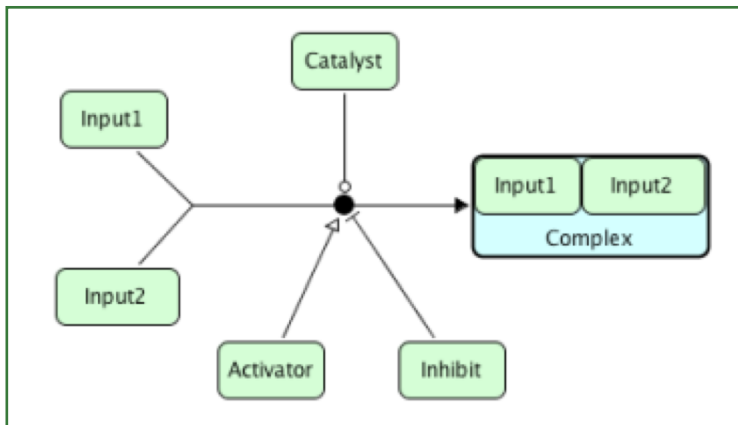
- Analyzing mutated genes in a network context:
 - reveals relationships among these genes.
 - can elucidate mechanism of action of drivers.
 - facilitates hypothesis generation on roles of these genes in disease phenotype.
- Network analysis reduces hundreds of mutated genes to < dozen mutated pathways.



What is a Functional Interaction?

- Convert reactions in pathways into pair-wise relationships
 - **Functional Interaction:** an interaction in which two proteins are involved in the same reaction as input, catalyst, activator and/or inhibitor, or as components in a complex

Reaction

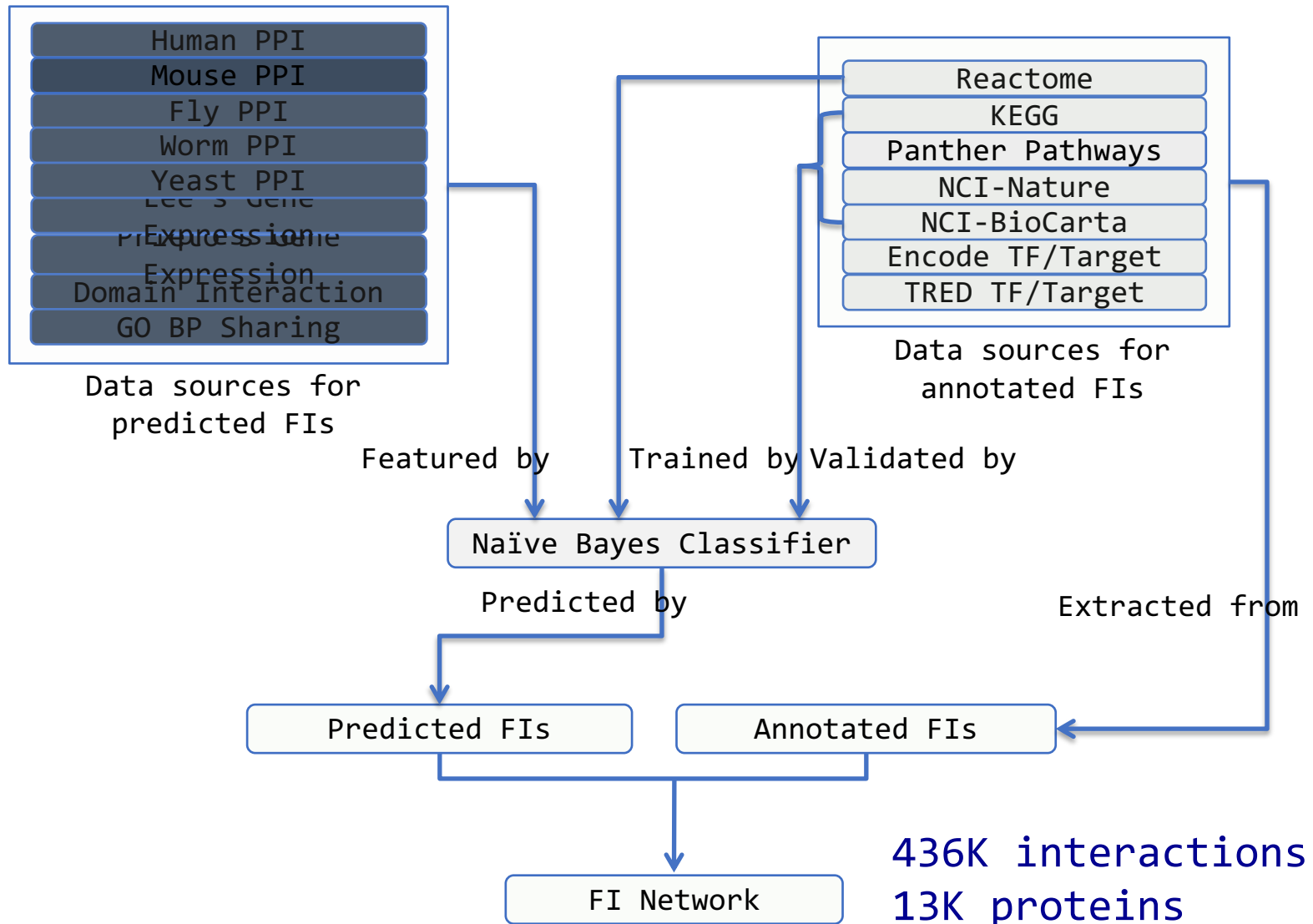


Functional Interactions

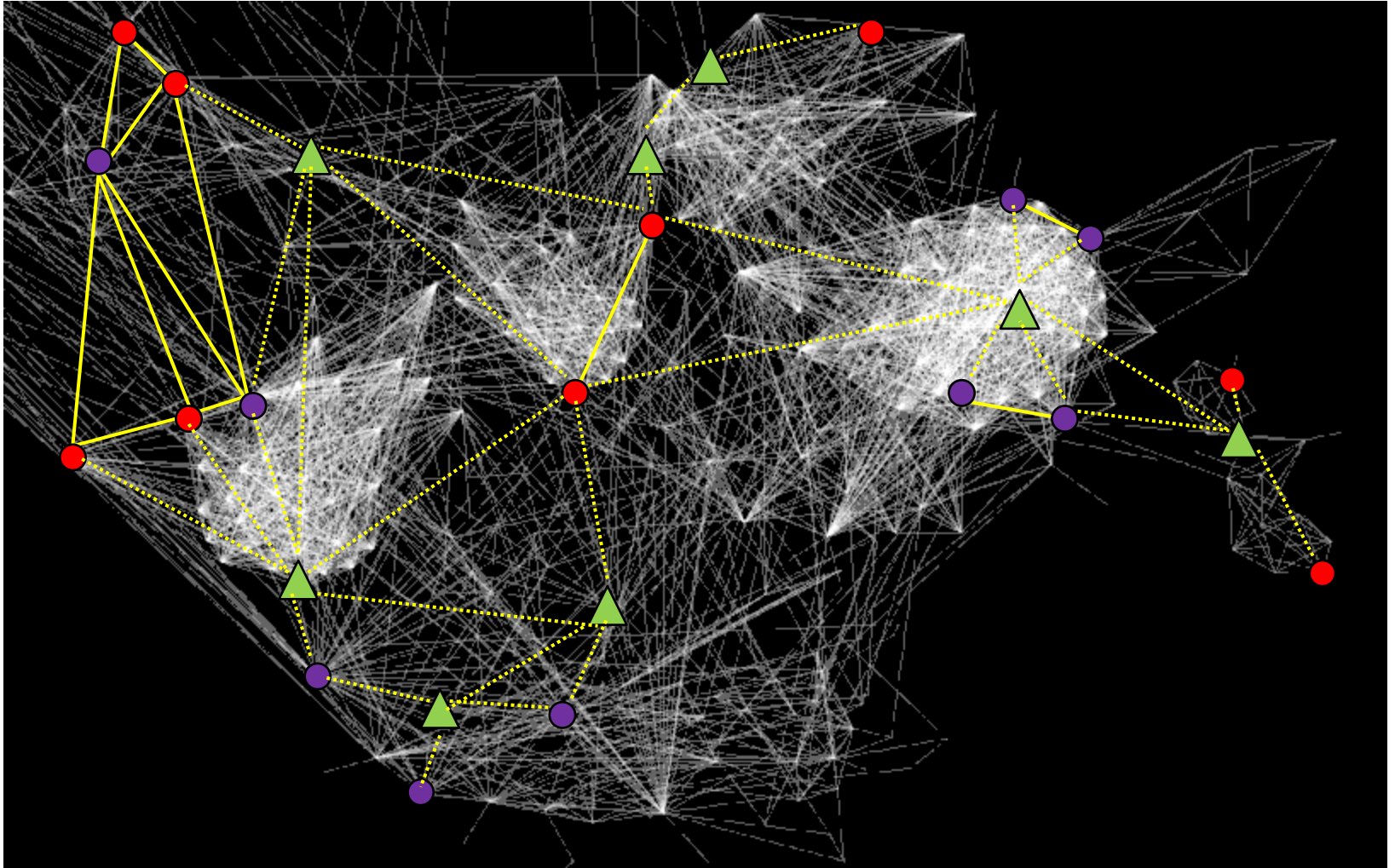
Input1-Input2, Input1-Catalyst,
Input1-Activator, Input1-Inhibitor, Input2-Catalyst,
Input2-Activator, Input2-Inhibitor,
Catalyst-Activator, Catalyst-Inhibitor, Activator-Inhibitor

[Wu et al. 2010 Genome Biology.](#)

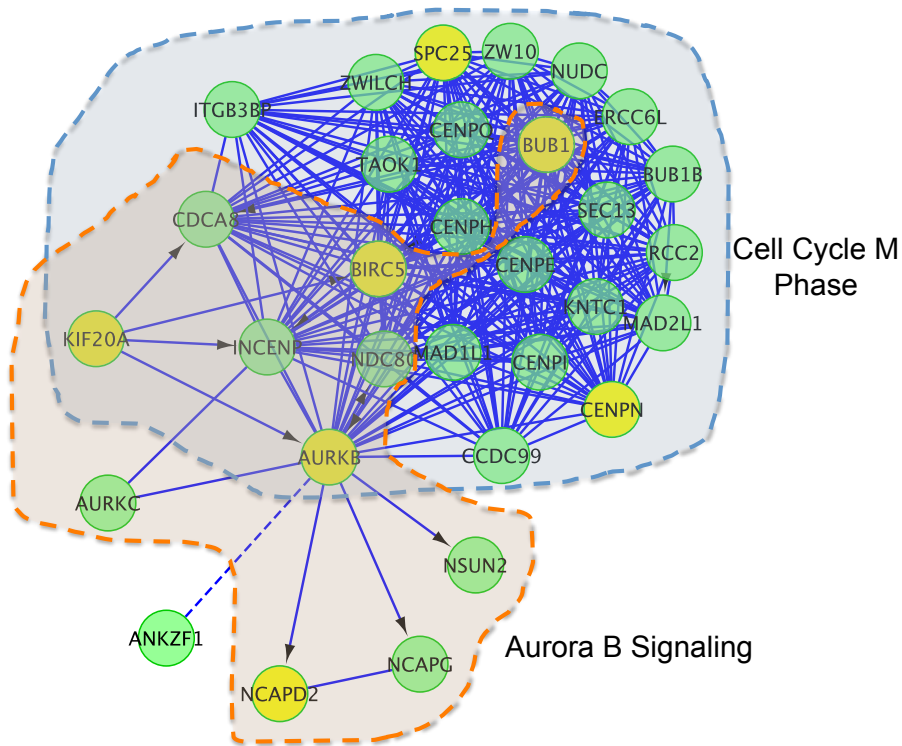
Construction of the FI Network



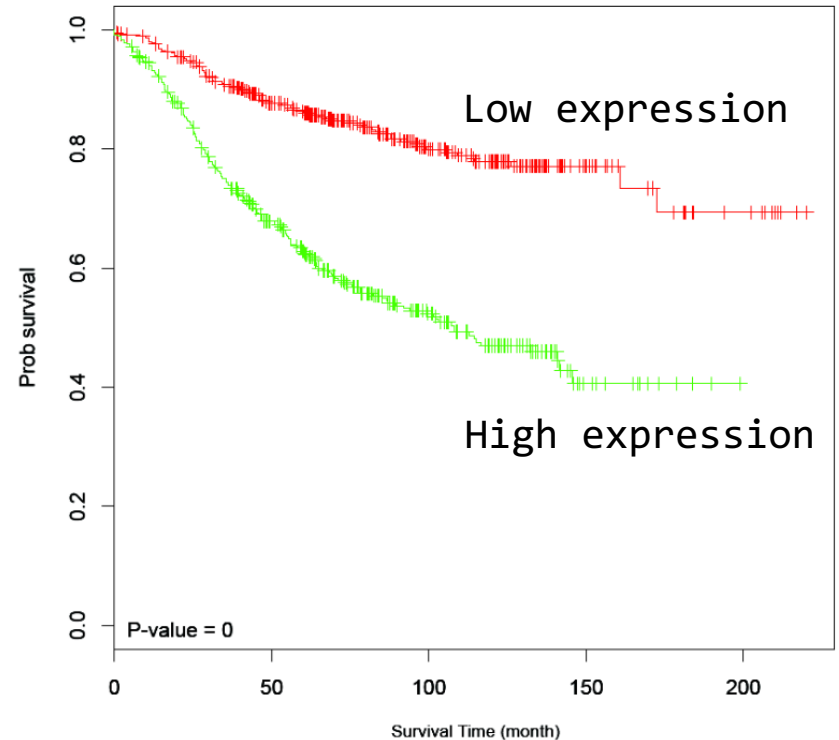
Projecting Experimental Data onto FI Network



Module-Based Prognostic Biomarker in ER+ Breast Cancer



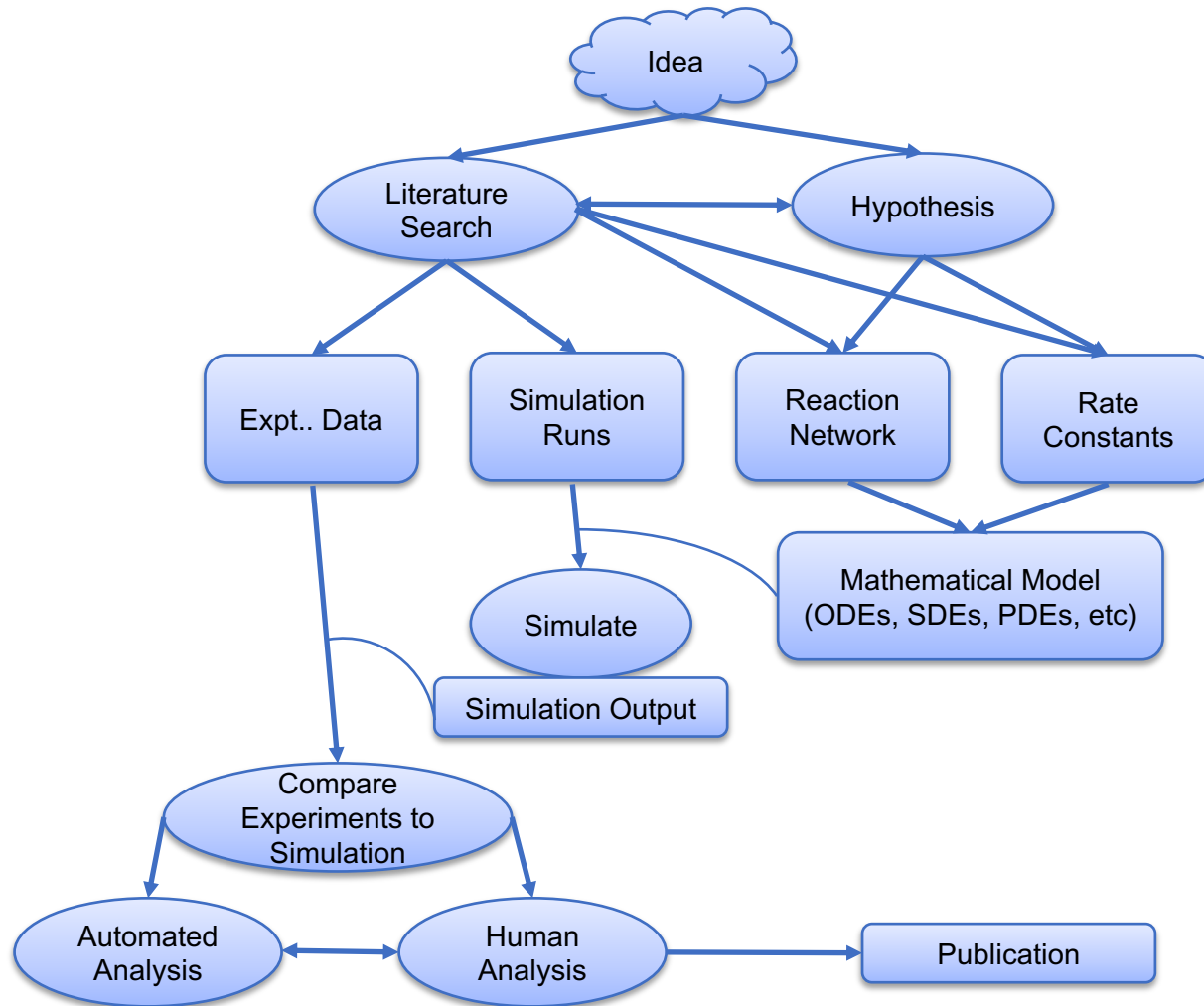
Measure levels of expression of the genes in this network module



Pathway modeling approaches

- Network based methods relate graph theory to discover and connect associations between nodes in the pathway, typical node and internodes interaction.
- Mathematical modeling learns and analyzes the underlying network by transforming the reactions and entities into matrix form.
- Preserve detailed biological relationships.
- Attempt to integrate multiple molecular alterations together to yield lists of altered pathway activities.
- Pathway modeling shades into Systems Biology

Modeling Lifecycle



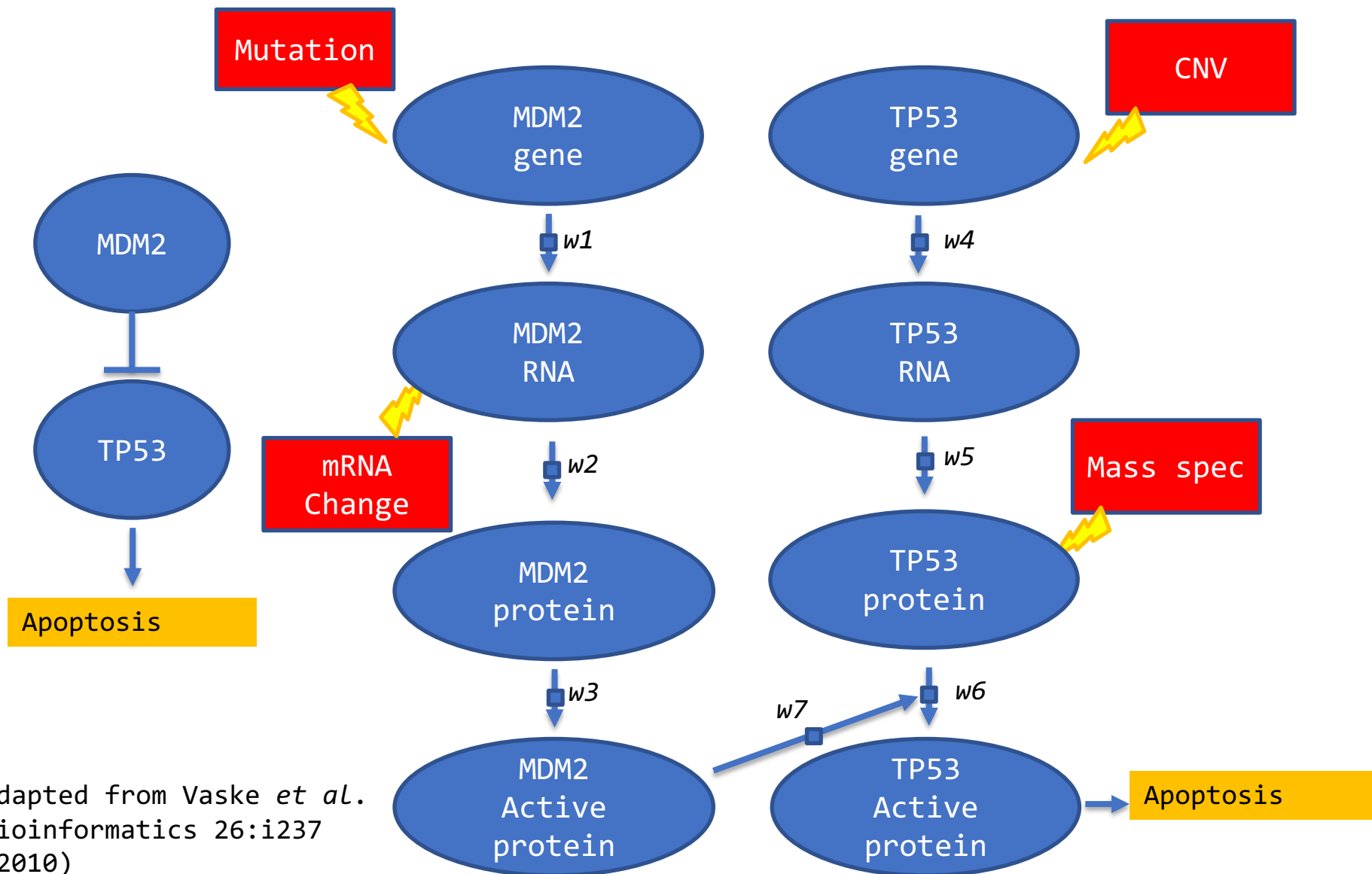
Types of Pathway-Based Modeling

- Partial differential equations/boolean models, e.g. CellNetAnalyzer
 - Mostly suited for biochemical systems (metabolomics)
- Network flow models, e.g. KinomeExplorer, NetPhorest, NetworKIN
 - Mostly suited for kinase cascades (phosphorylation info)
- Transcriptional regulatory network-based reconstruction methods, e.g. ARACNe (expression arrays)
- Probabilistic graph models (PGMs), e.g. PARADIGM
 - Most general form of pathway modeling for cancer analysis at this time.

Probabilistic Graphical Model (PGM) based Pathway Analysis

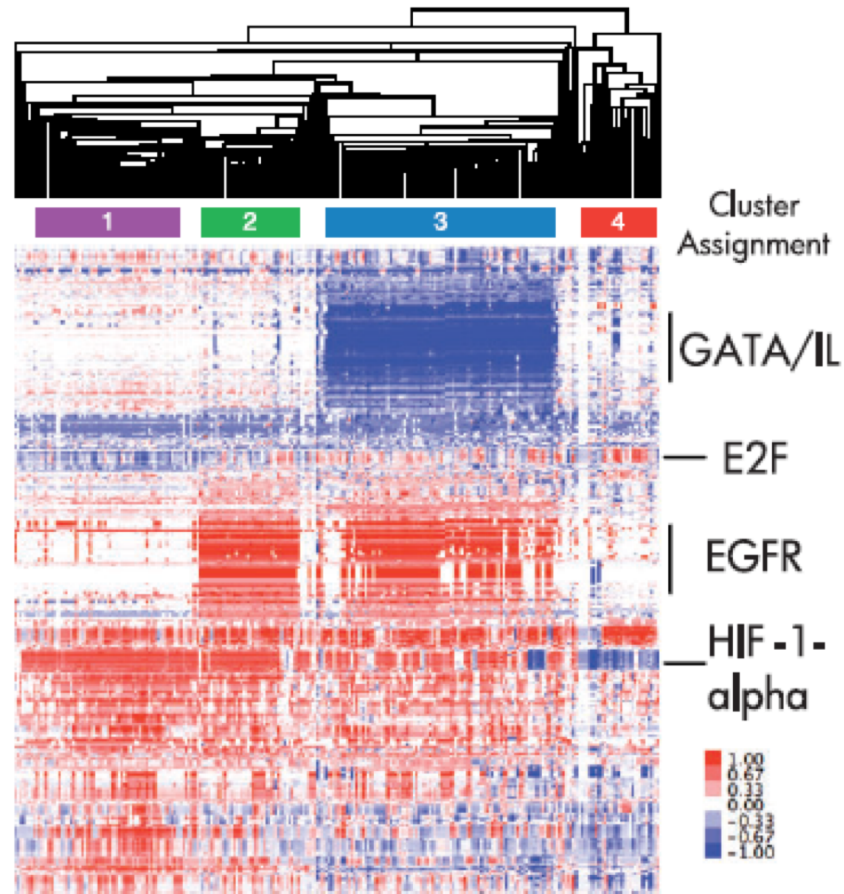
- Attempt to integrate multiple molecular alterations together to yield lists of altered pathway activities.
 - Many omics data types available for one single patient
 - CNV, gene expression, methylation, somatic mutations, etc.
- Pathway and Network-based Simulation
 - Find significantly impacted pathways for diseases
 - Link pathway activities to patient phenotypes
 - Predict drug effects: one drug or multiple drugs together?

PARADIGM



Adapted from Vaske *et al.*
Bioinformatics 26:i237
(2010)

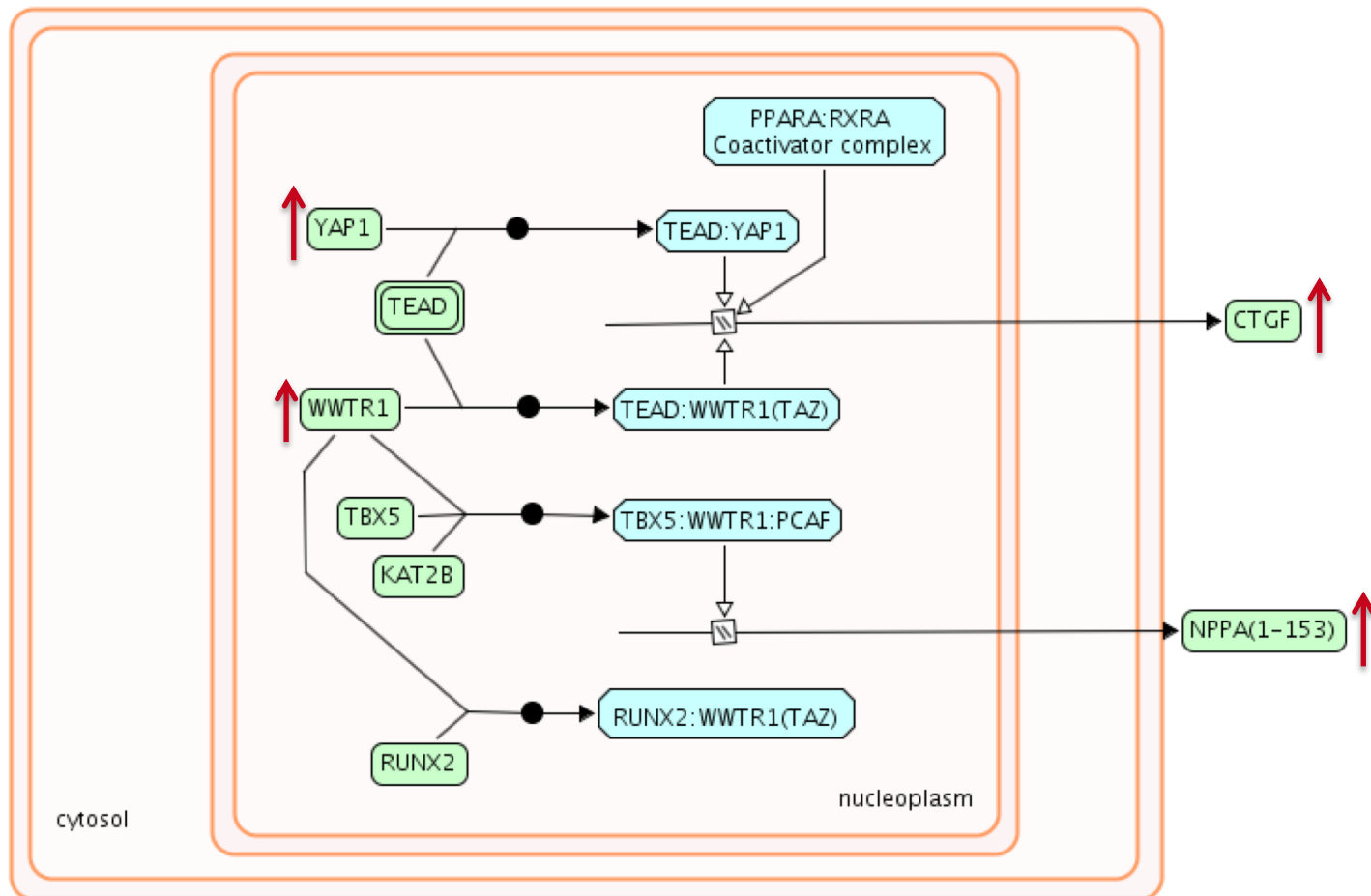
PARADIGM Applied to GBM Data



Vaske *et al.*
Bioinformatics 26:i237
(2010)

Fig. 8. Clustering of IPAs for TCGA GBM. Each column corresponds to a single sample, and each row to a biomolecular entity. Color bars beneath the hierarchical clustering tree denote clusters used for Figure 9.

Probabilistic Graphical Models (PGMs) for Reactome Pathways



Pathway/Network Database URLs

- BioGRID
 - [http:// www.thebiogrid.org](http://www.thebiogrid.org)
- IntAct
 - <http://www.ebi.ac.uk/intact/>
- KEGG
 - [http:// www.genome.jp/kegg](http://www.genome.jp/kegg)
- MINT
 - <http://mint.bio.uniroma2.it>
- Reactome
 - [http:// www.reactome.org](http://www.reactome.org)
- Pathway Commons
 - <http://www.pathwaycommons.org>
- WikiPathways
 - <http://wikipathways.org>

De novo network construction & clustering

- GeneMANIA
 - <http://www.genemania.org>
- HotNet
 - <http://compbio.cs.brown.edu/projects/hotnet/>
- HyperModules
 - <http://apps.cytoscape.org/apps/hypermodules>
- Reactome Cytoscape FI App
 - <http://apps.cytoscape.org/apps/reactomefis>

Pathway Modeling

- CellNetAnalyzer
 - <https://www2.mpi-magdeburg.mpg.de/projects/cna/cna.html>
- KinomeExplorer/NetPhorest/Networkin
 - <http://KinomeExplorer.info>
 - <http://netphorest.info>
 - <http://networkin.info>
- ARACNe
 - <http://califano.c2b2.columbia.edu/aracne>
- PARADIGM
 - <http://paradigm.five3genomics.com/>

We are on a Coffee Break & Networking Session

compute | calcul
canada | canada

Workshop Sponsors:



Canadian Centre for
Computational
Genomics

MiCM McGill initiative in
Computational Medicine