

A complex network diagram composed of numerous thin, light-blue curved lines connecting small, dark-blue circular nodes. The nodes are densely packed in several clusters, with one prominent central cluster and smaller groups extending towards the edges. The overall effect is a sense of dynamic flow and interconnectedness.

# Pathways and Networks

An Overview

Contact: [help@reactome.org](mailto:help@reactome.org)



reactome

# Learning Objectives of Module

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



# What is Pathway/Network Analysis?

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



# Why Pathway/Network Analysis?

- Dramatic data size reduction: 1000's of genes => dozens of pathways.
- Increase statistical power by reducing multiple hypotheses.
- Find meaning in the “long tail” of rare cancer mutations.
- Tell biological stories:
  - Identifying hidden patterns in gene lists.
  - Creating mechanistic models to explain experimental observations.
  - Predicting the function of unannotated genes.
  - Establishing the framework for quantitative modeling.
  - Assisting in the development of molecular signatures.

# Why Pathway Analysis?

## Mutational landscape and significance across 12 major cancer types

Cyriac Kandoth<sup>1\*</sup>, Michael D. McLellan<sup>1\*</sup>, Fabio Vandin<sup>2</sup>, Kai Ye<sup>1,3</sup>, Beifang Niu<sup>1</sup>, Charles Lu<sup>1</sup>, Mingchao Xie<sup>1</sup>, Qunyuan Zhang<sup>1,3</sup>, Joshua F. McMichael<sup>1</sup>, Matthew A. Wyczalkowski<sup>1</sup>, Mark D. M. Leiserson<sup>2</sup>, Christopher A. Miller<sup>1</sup>, John S. Welch<sup>4,5</sup>, Matthew J. Walter<sup>4,5</sup>, Michael C. Wendl<sup>1,3,6</sup>, Timothy J. Ley<sup>1,3,4,5</sup>, Richard K. Wilson<sup>1,3,5</sup>, Benjamin J. Raphael<sup>2</sup> & Li Ding<sup>1,3,4,5</sup>

### 127 Cancer Driver Genes

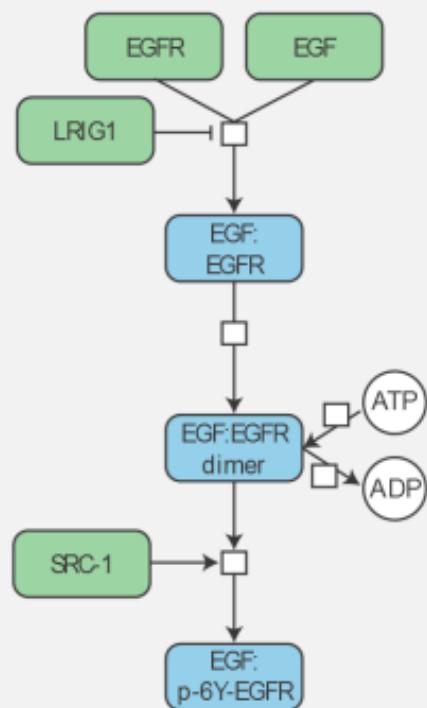
ACVR1B, ACVR2A, AJUBA, AKT1, APC, AR, ARHGAP35, ARID1A, ARID5B, ASXL1, ATM, ATR, ATRX, AXIN2, B4GALT3, BAP1, BRAF, BRCA1, BRCA2, CBFB, CCND1, CDH1, CDK12, CDKN1A, CDKN1B, CDKN2A, CDKN2C, CEBPA, CHEK2, CRIPAK, CTCF, CTNNB1, DNMT3A, EGFR, EGR3, EIF4A2, ELF3, EP300, EPHA3, EPHB6, EPPK1, ERBB4, ERCC2, EZH2, 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

Nature 502 (2013): 333-339

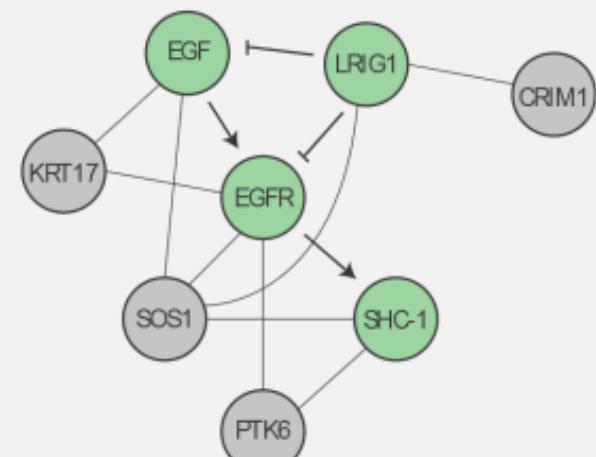


# Pathways vs Networks

EGFR-centered  
Pathway



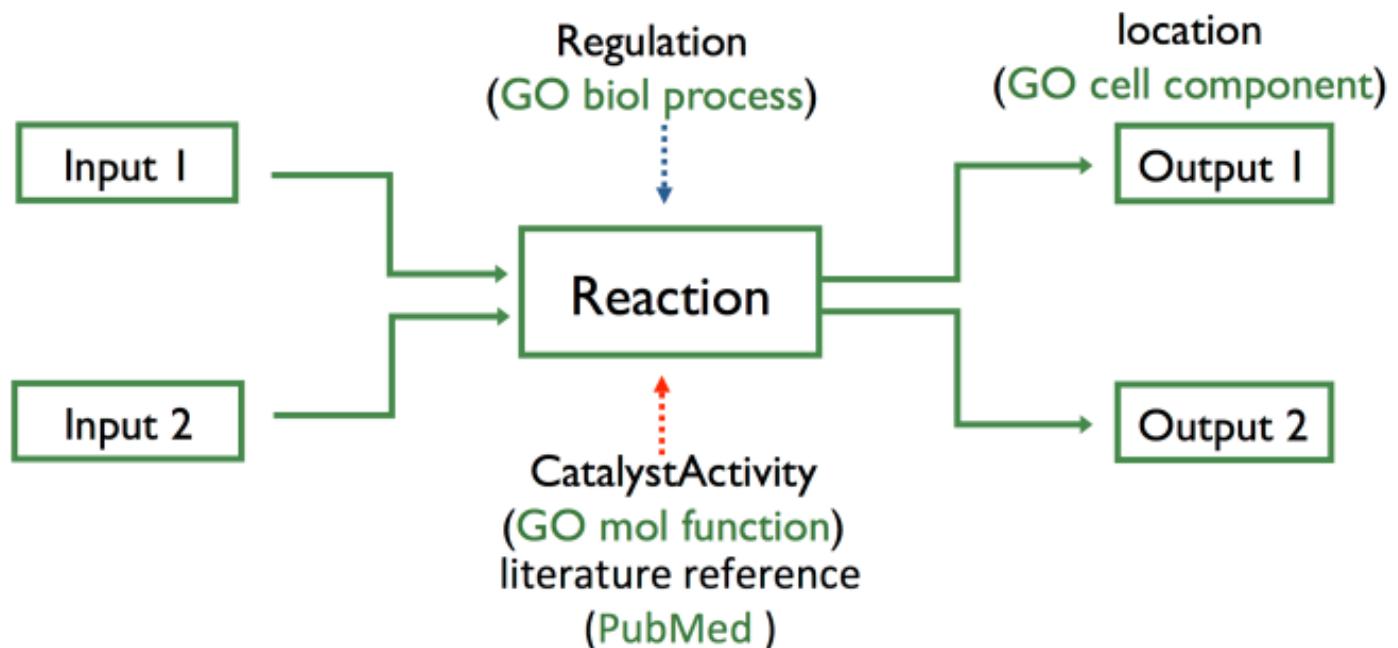
EGFR-centered  
Network



# Reaction-Network Databases

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

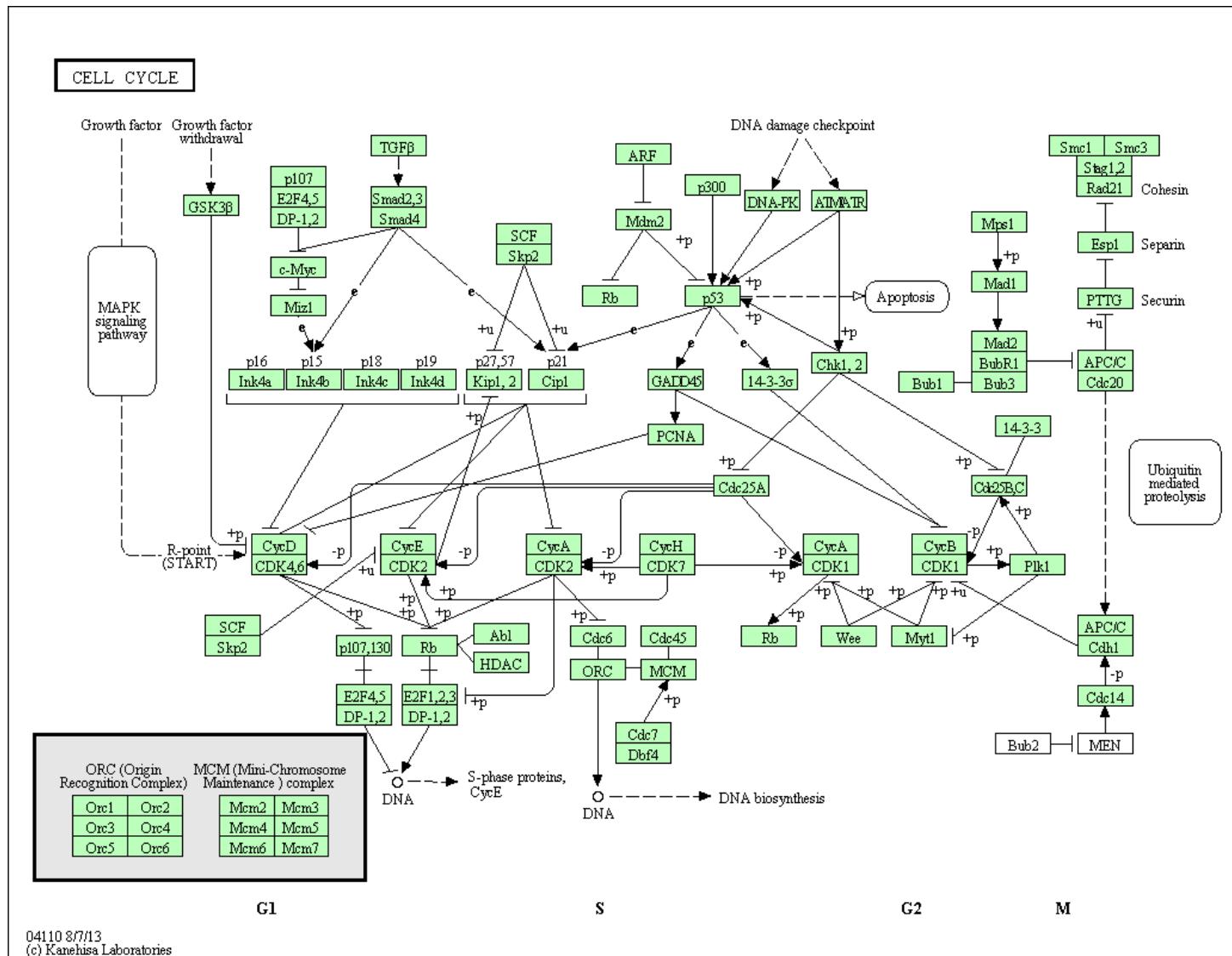
protein (UniProt)  
molecule (ChEBI)  
complex (GO/PRO)  
ncRNA (miRBase)  
disease variants (UniProt)  
therapeutics (ChEBI)



# KEGG

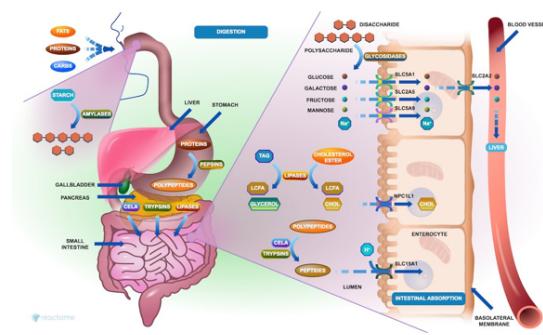
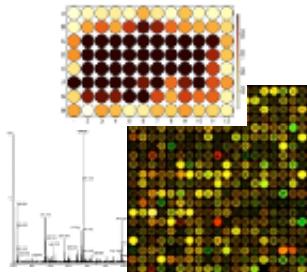
- KEGG is a collection of biological information compiled from published material → **curated database**.
- Includes information on genes, proteins, metabolic pathways, molecular interactions, and biochemical reactions associated with specific organisms
- Provides a relationship (map) for how these components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

# KEGG Pathway

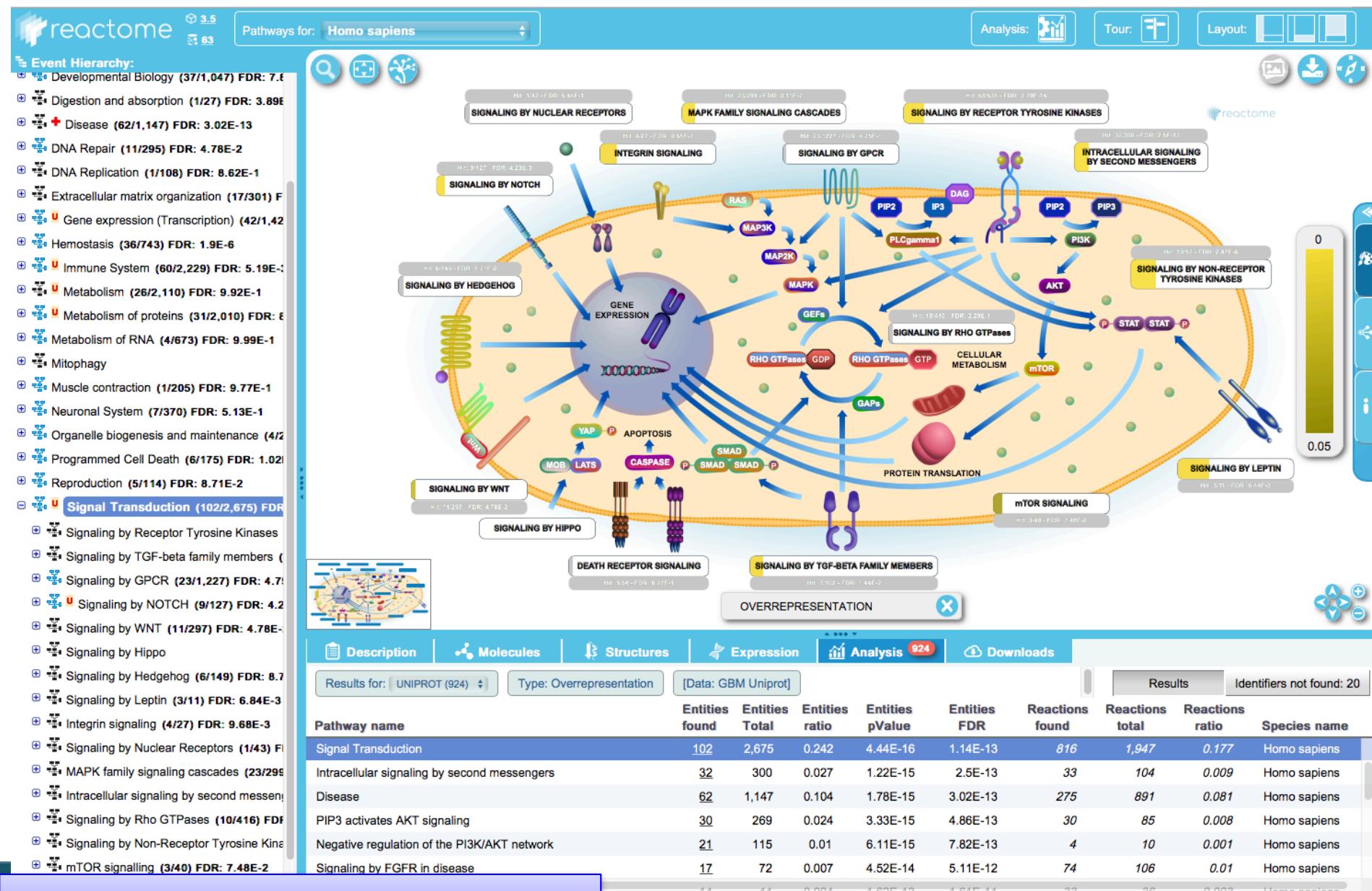


# Reactome

- 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

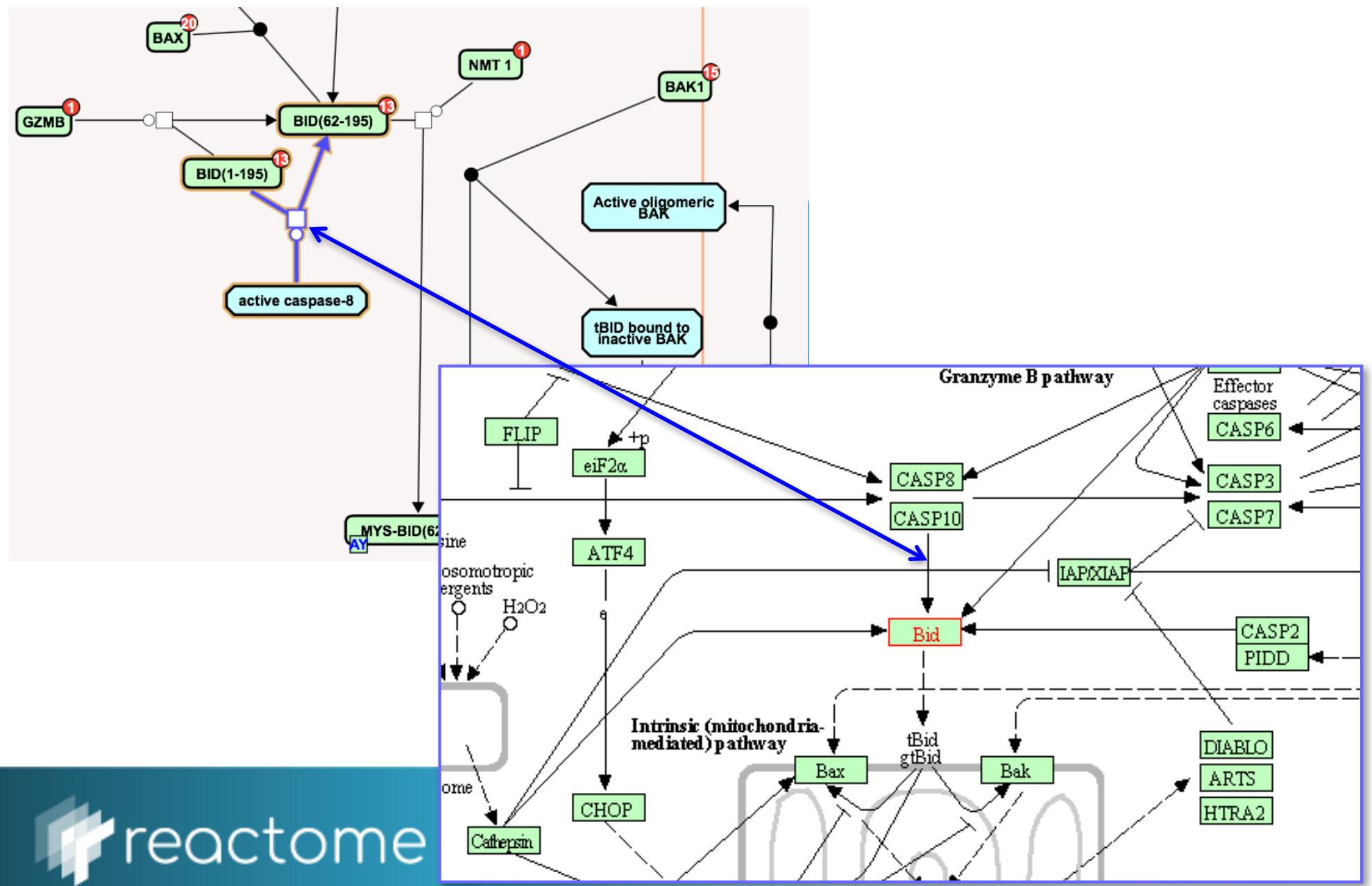


# Pathway Browser: Pathway Enrichment Analysis



127 Cancer Driver Genes

# Reactome vs KEGG



# Pathway Commons

 **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

Names or gene IDs (e.g. 'glycolysis', 'TP53')

**PCViz**

Get details about genes and their interactions

Gene IDs (e.g. 'MDM2 TP53')

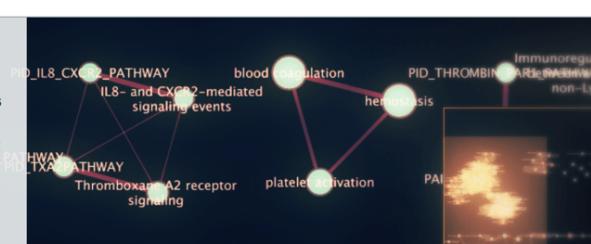
---

## Training

**Guide**

LPA receptor-mediated events PID\_IL8\_CXCL8\_PATHWAY IL8- and CXCL2-mediated signaling events blood coagulation PID\_THROMBIN immunoregulatory pathways non-Lytic Endocrinins

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.

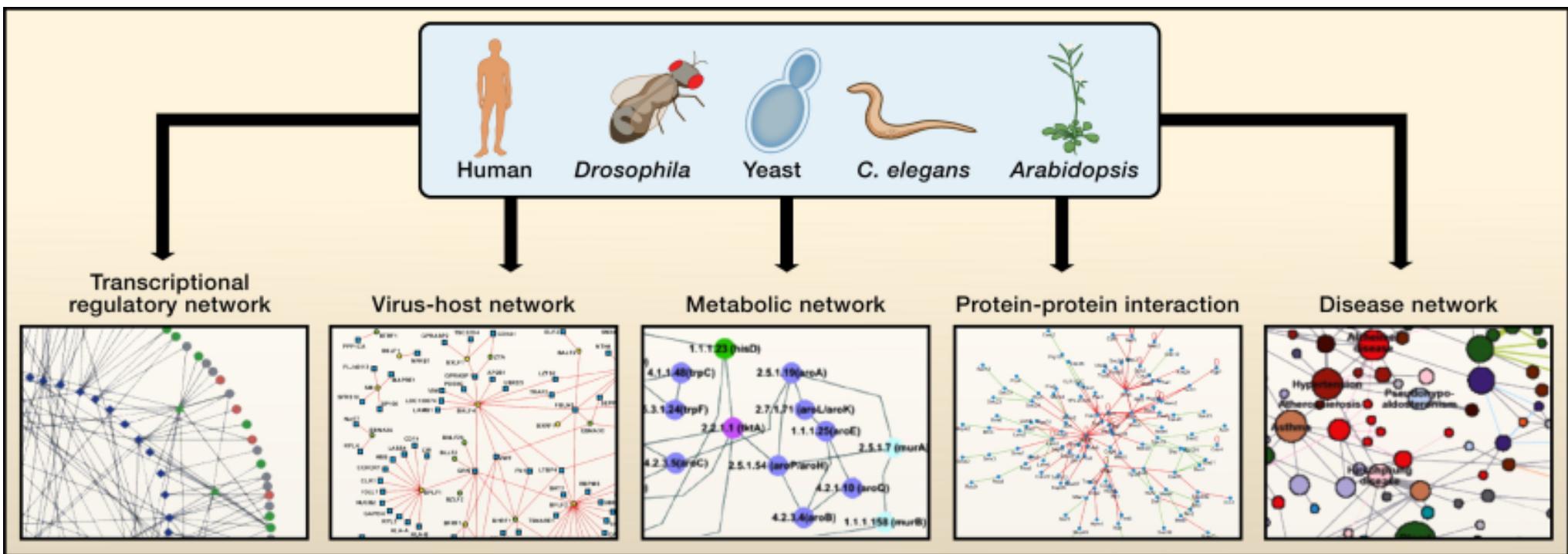


# What is a Interaction Network?

- An Interaction Network is a collection of:
  - Nodes (or vertices).
  - Edges connecting nodes (directed or undirected, weighted, multiple edges, self-edges).
- Nodes can represent proteins, genes, metabolites, or groups of these (e.g. complexes) - any sort of object.
- Edges can be either physical or functional interactions, activators, regulators, reactions - any sort of relations.



# Types of Interactions Networks



# Interaction Databases

- Can be built automatically or via curation.
- More extensive coverage of biological systems.
- Relationships and underlying evidence more tentative.
- Popular sources of curated human networks:
  - BioGRID – Curated physical and genetic interactions from literature; 65K interactors & 1.1M interactions.
  - IntAct – Curated interactions from literature; 95K interactors & 694K interactions.

# IntAct

EMBL-EBI

IntAct

Services Research Training About us

TP53  
Examples: BRCA2, Q06609, dmc1, 10631611

Search Advanced

Home Advanced Search About Resources Download Feedback

IntAct > IntAct Search Results

+ Show more data from EMBL-EBI

## 11,128 binary interactions found for search term TP53

Interactions (11128) Interactors Interaction Details Graph

Filter out the spoke expanded co-complexes (3924)  
Include 23 negative interactions matching the query (pink row, not exported in MITAB 2.5, RDF, Biopax, XGMML)

Your query also matches 28,035 interaction evidences from 10 other databases.  
Your query also matches 9,109 interaction evidences from 1 other IMEx databases.

What is this view?

Dts	Molecule 'A'	Links 'A'	Molecule 'B'	Links 'B'	Interaction Detection Method	Interaction AC	Source Database
1	TP53	P04637 EBI-366083	MDM2	Q00987 EBI-389668	anti bait coimmunoprecipitation	EBI-7316419 MINT-4545014 imex : IM-11374-7	MINT
2					ubiquitinase assay	EBI-9350109 imex : IM-22307-12	IntAct
3					ubiquitinase assay	EBI-9350102 imex : IM-22307-11	IntAct
4					enzyme linked immunosorbent assay	EBI-9074450 imex : IM-21984-2	IntAct
5					enzyme linked immunosorbent assay	EBI-9074437 imex : IM-21984-1	IntAct
6					proximity ligation assay	EBI-6593297 imex : IM-18707-329	IntAct
7					anti tag coimmunoprecipitation	EBI-7902921 MINT-8415355 imex : IM-20828-4	MINT
8					anti tag coimmunoprecipitation	EBI-7156209 MINT-4054304 imex : IM-11399-3	MINT

[www.ebi.ac.uk/intact/](http://www.ebi.ac.uk/intact/)



# Open Data Exchange



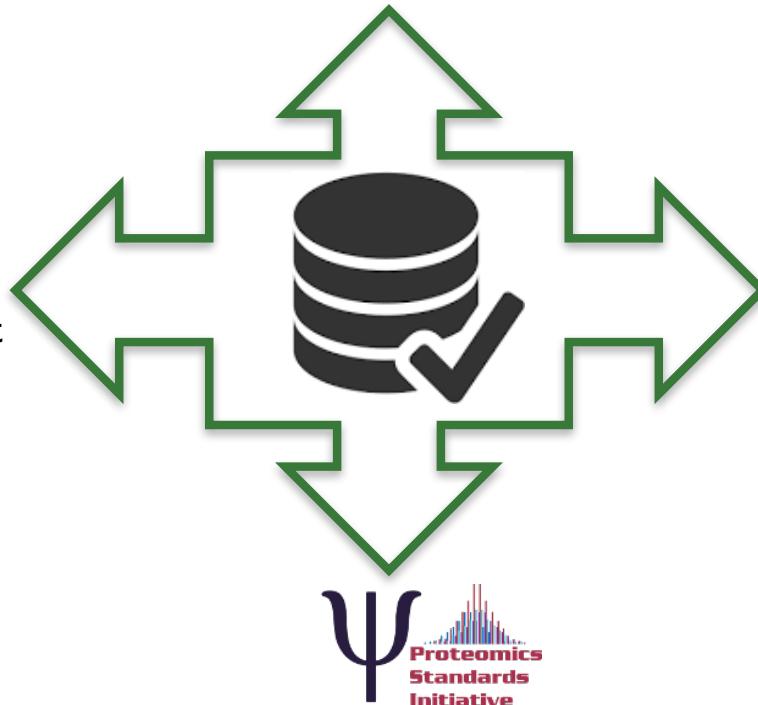
Standard graphical languages for  
representing biological processes  
and interactions



SBML level 3.  
Open access interchange format  
for computer models of  
biochemical pathways, reactions  
and networks.



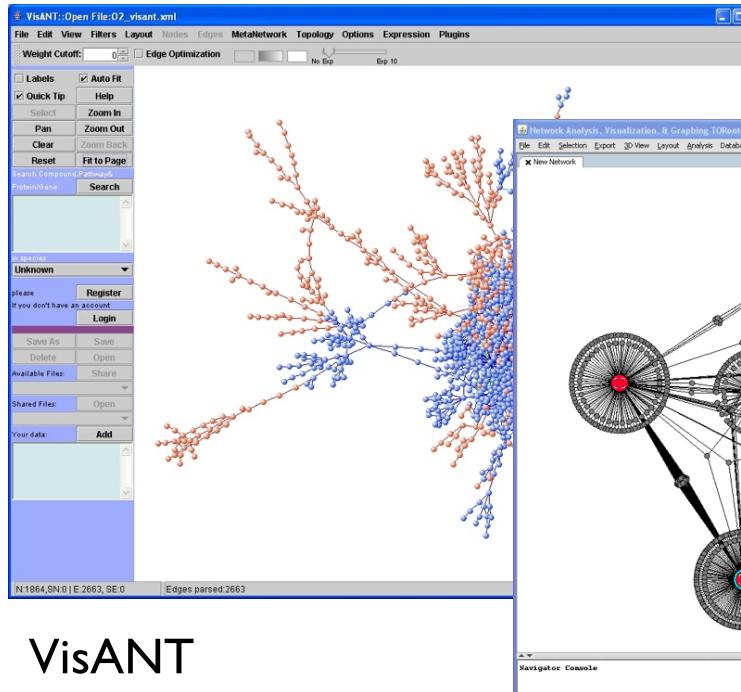
BioPAX level 3.  
Standard language that aims to  
enable integration, exchange,  
visualization and analysis of  
biological pathway data.



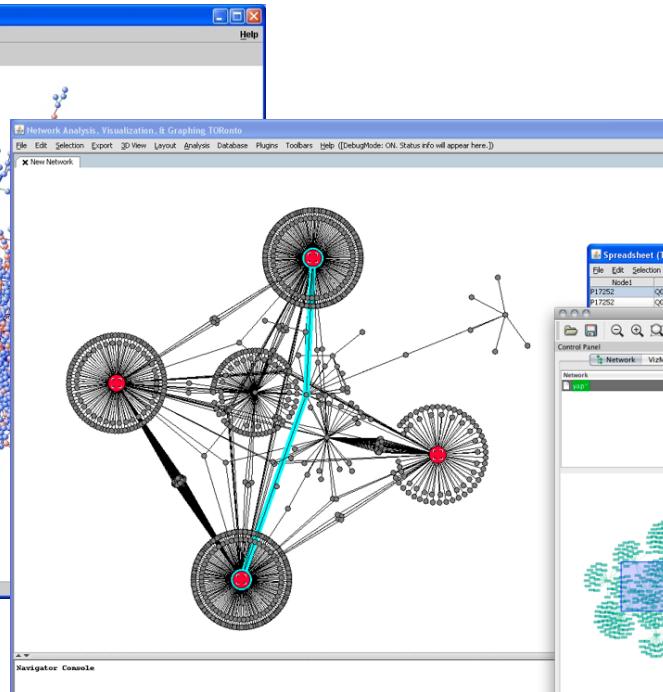
PSICQUIC is an effort to standardize the  
access to molecular interaction databases.  
PSI-MITAB is the data exchange format.



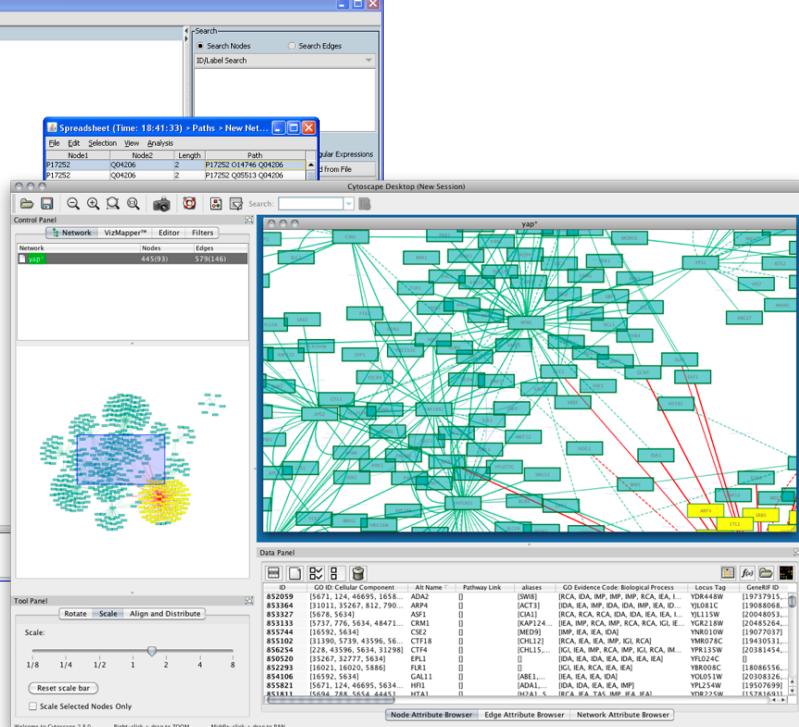
# Visualization and Analysis Tools for Biological Networks



VisANT



NAVIGATOR



Cytoscape



EMBL-EBI



NYU Langone  
MEDICAL CENTER

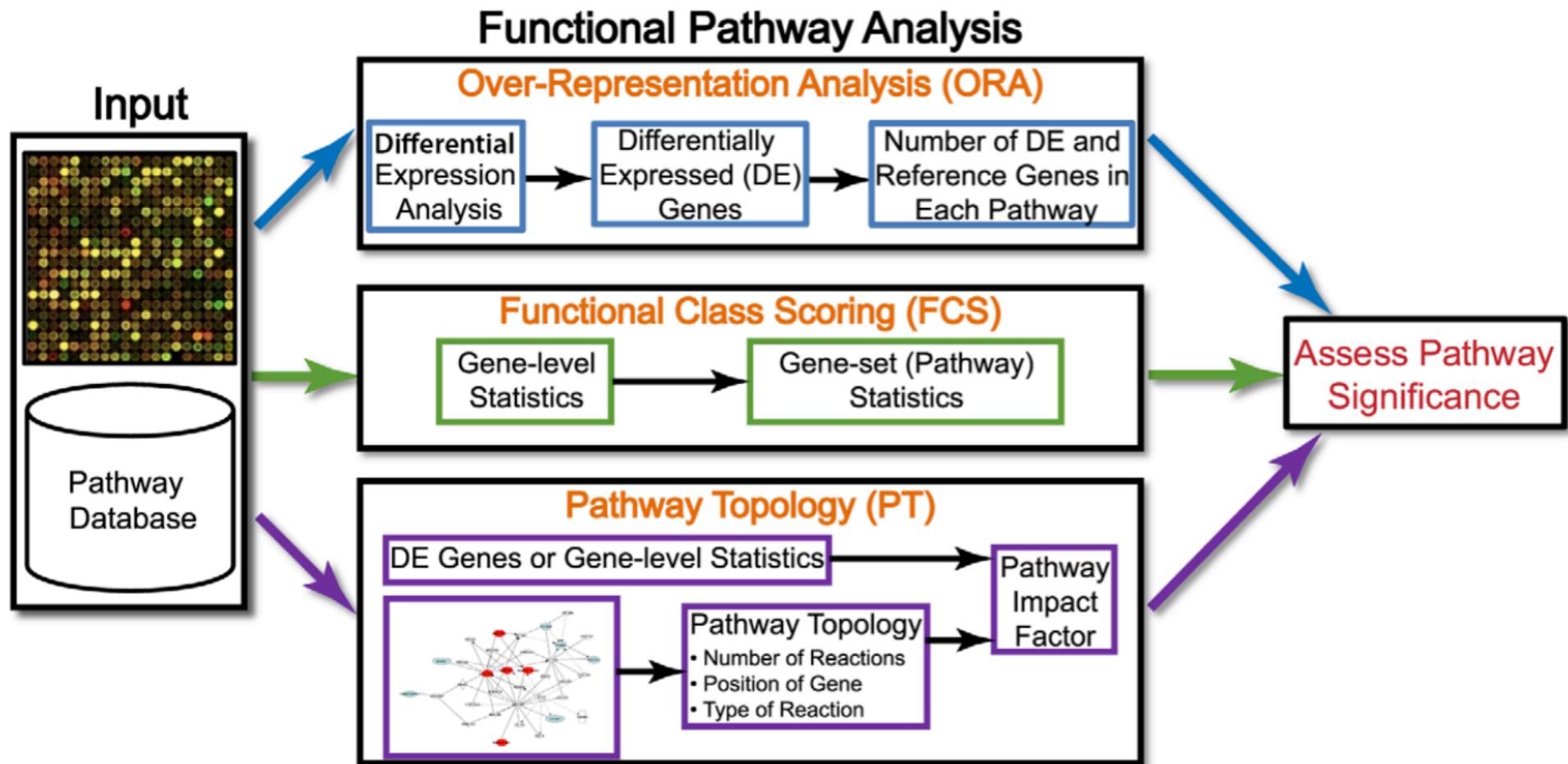


GHIU

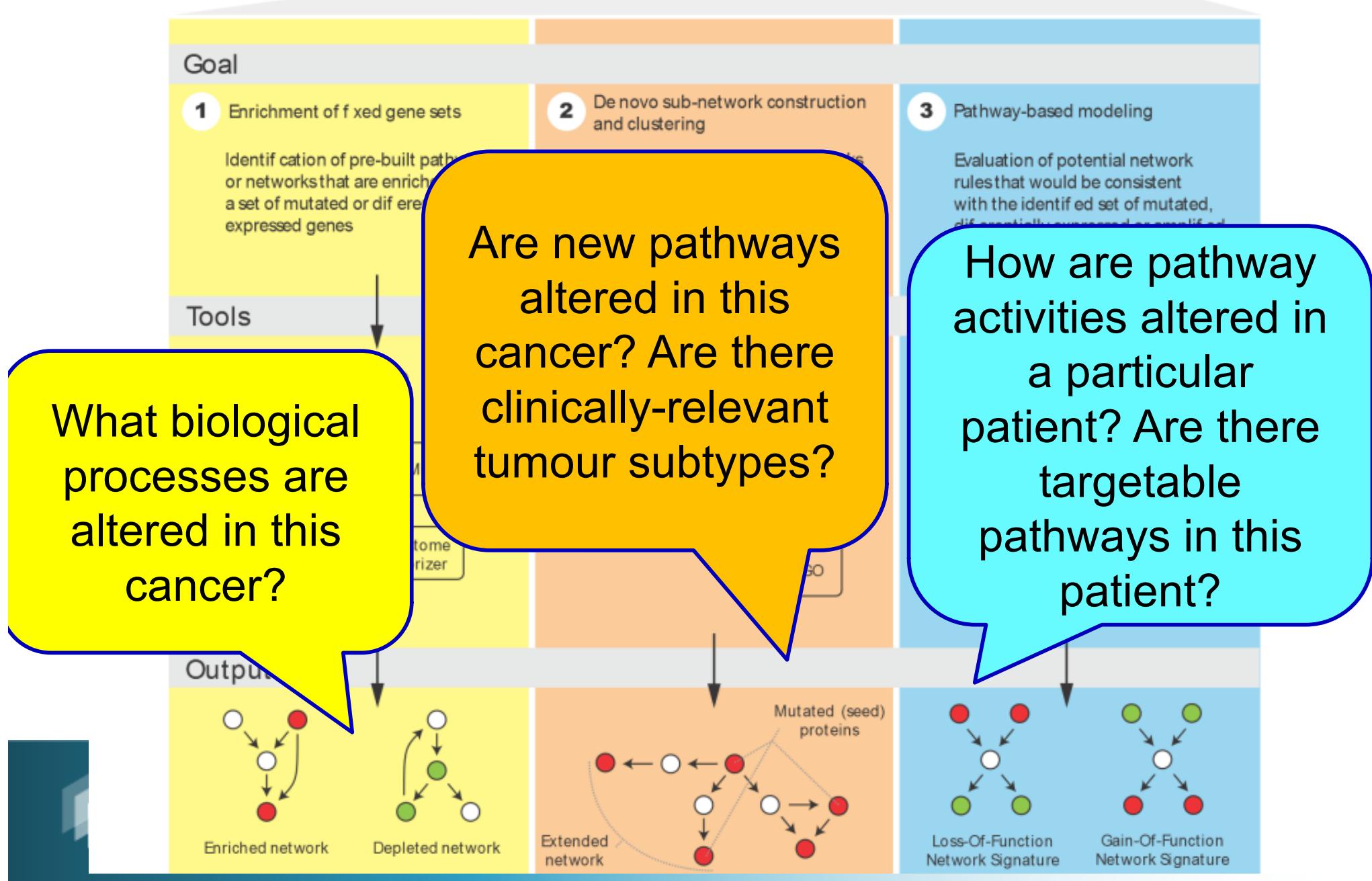


OICR  
Ontario Institute  
for Cancer Research

# Pathway/Network Analysis Workflow



# Types of Pathway/Network Analysis



# Issues in Pathway-based Analysis

- How to handle the pathway hierarchical organization?
  - Flatten pathways organized in a hierarchy into a systems-wide network
- How to handle pathway cross-talks?
  - Shared genes/proteins will be listed once in a network
  - Interactions causing cross-talks are displayed in the same network
- How to integrate omics data types available for one single patient
  - CNV, gene expression, methylation, somatic mutations, etc.
- Pathway and network-based Simulation
  - How to use topological structures?
  - Predict drug effects: one drug or multiple drugs together?

# Network-based Data Analysis

- Based on systems-wide biological networks
  - Covering the majority of human genes
  - Usually protein-protein interaction networks
- Modules (or clusters)-based network patterns
  - Pathway annotations via enrichment analysis
  - Gene signature or biomarker discovery
- Disease genes discovery
  - Cancer drivers
  - Disease modules



## 2) De Novo Subnetwork Construction & Clustering

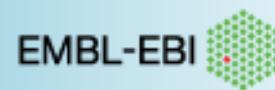
- Apply list of altered {genes,proteins,RNAs} to a biological network.
- Identify “topologically unlikely” configurations.
  - E.g. a subset of the altered genes are closer to each other on the network than you would expect by chance.
- Extract clusters of these unlikely configurations.
- Annotate the clusters.

# Network clustering

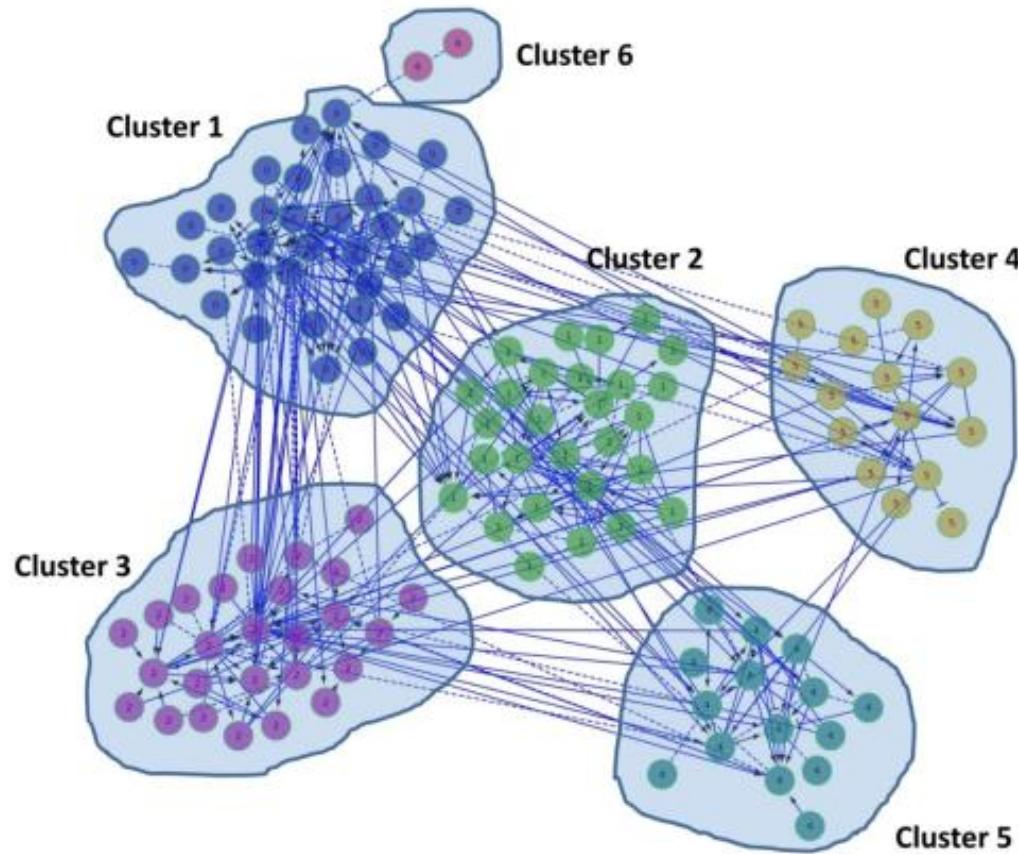
- Clustering can be defined as the process of grouping objects into sets called clusters (communities or modules), so that each cluster consists of elements that are similar in some ways.
- Network clustering algorithm is looking for sets of nodes [proteins] that are joined together in tightly knit groups.
- Cluster detection in large networks is very useful as highly connected proteins are often sharing similar functionality.

# Popular Network Clustering Algorithms

- Girvin-Newman
  - a hierarchical method used to detect communities by progressively removing edges from the original network
- Markov Cluster Algorithm
  - a fast and scalable unsupervised cluster algorithm for graphs based on simulation of (stochastic) flow in graphs
- HotNet
  - Finds “hot” clusters based on propagation of heat across metallic lattice.
  - Avoids ascertainment bias on unusually well-annotated genes.
- HyperModules Cytoscape App
  - Find network clusters that correlate with clinical characteristics.
- Reactome FI Network & ReactomeFIViz app
  - Offers multiple clustering and correlation algorithms (including HotNet, PARADIGM and survival correlation analysis)



# Typical output of a network clustering algorithm



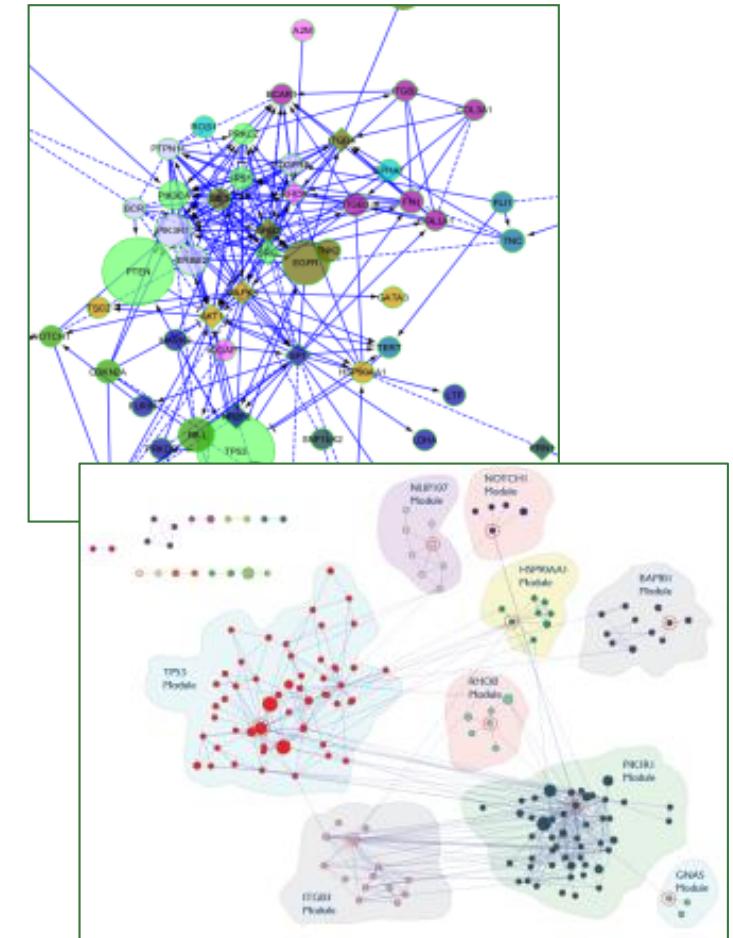
This hypothetical subnetwork was decomposed onto 6 clusters.

Different clusters are marked with different colors. Cluster 6 contains only 2 elements and could be ignored in the further investigations.

Clusters are mutually exclusive meaning that nodes are not shared between the clusters.

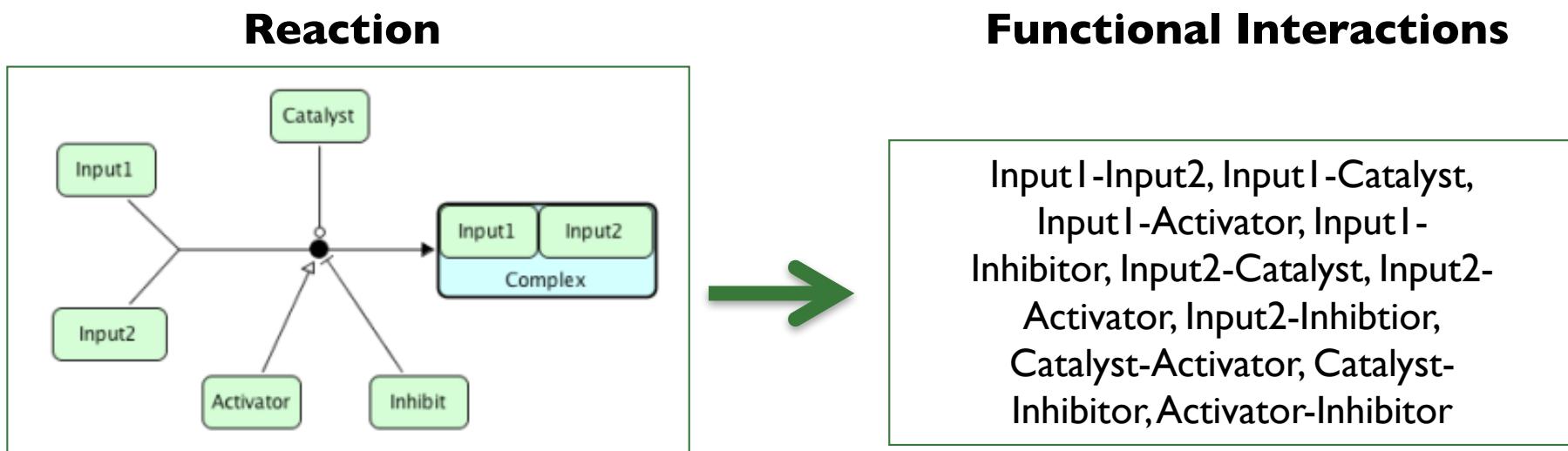
# Reactome Functional Interaction (FI) Network and ReactomeFlViz app

- No single mutated gene is necessary and sufficient to cause cancer.
  - Typically one or two common mutations (e.g. TP53) plus rare mutations.
- 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?

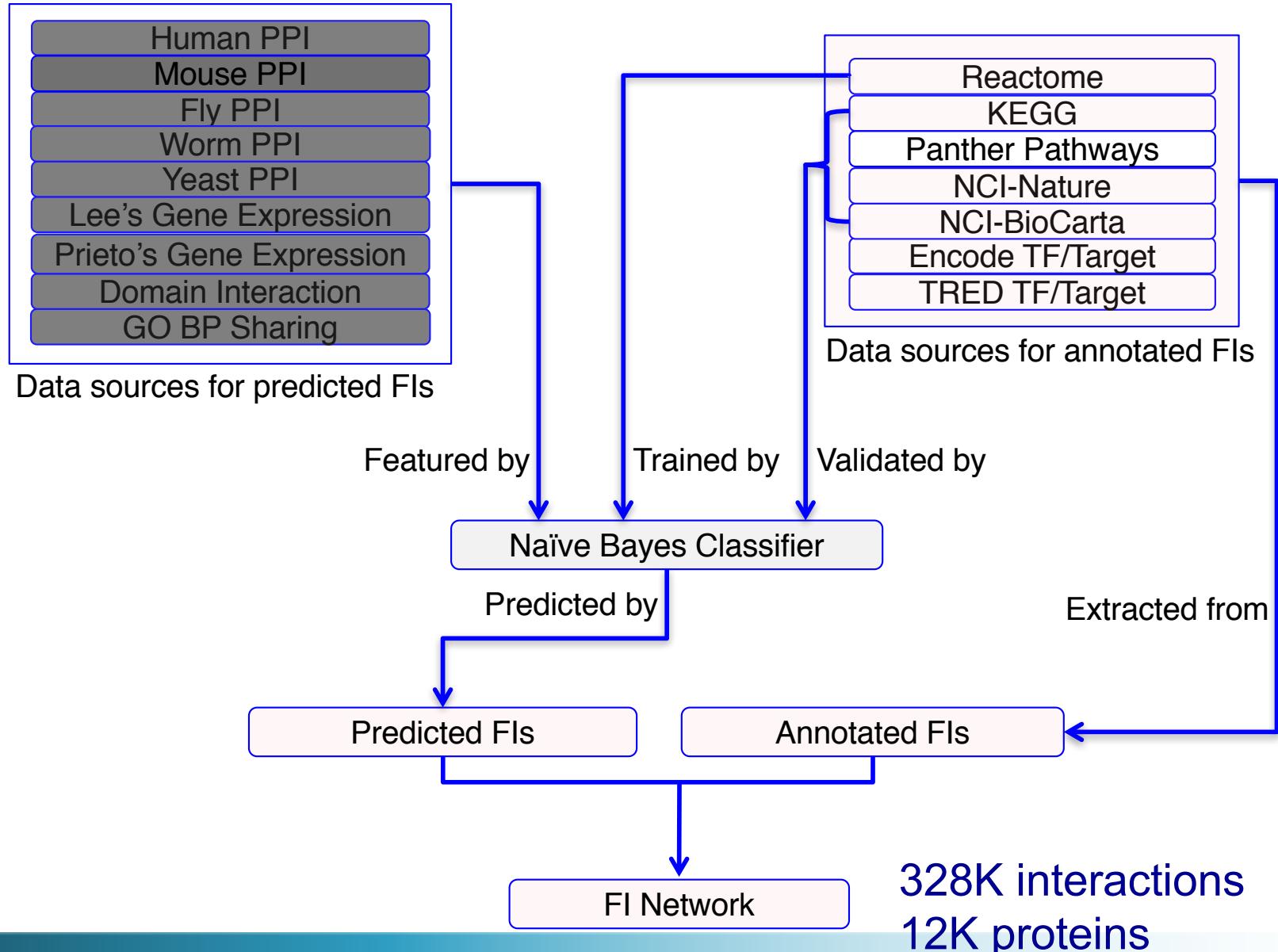
- Convert reactions in pathways into pair-wise relationships
  - **Functional Interaction:** an interaction in which two proteins are involved in the same reaction as input, catalyst, activator and/or inhibitor, or as components in a complex



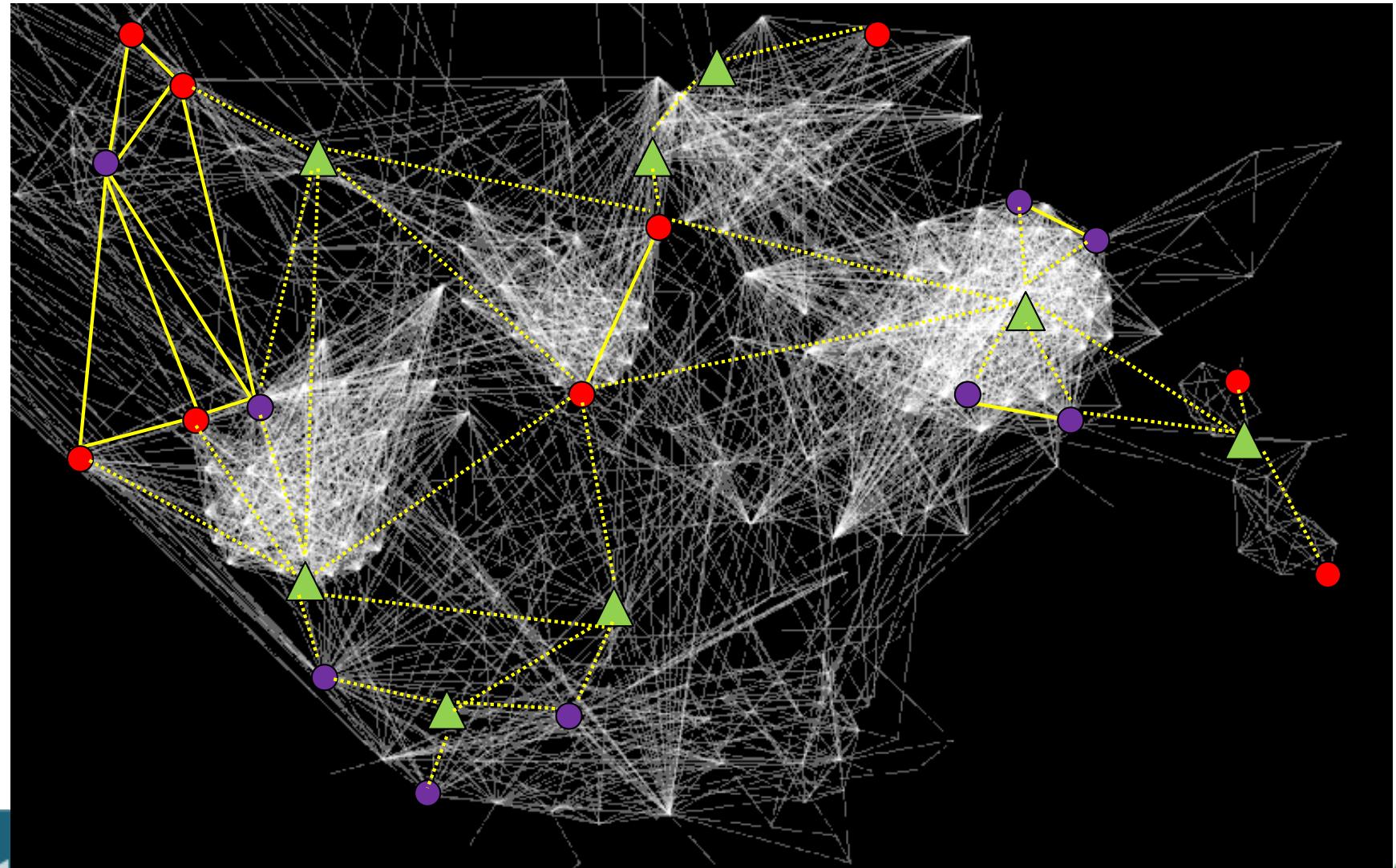
Method and practical application: A human functional protein interaction network and its application to cancer data analysis, [Wu et al. 2010 Genome Biology](#).



# Construction of the FI Network



# Projecting Experimental Data onto FI Network



reactome

EMBL-EBI



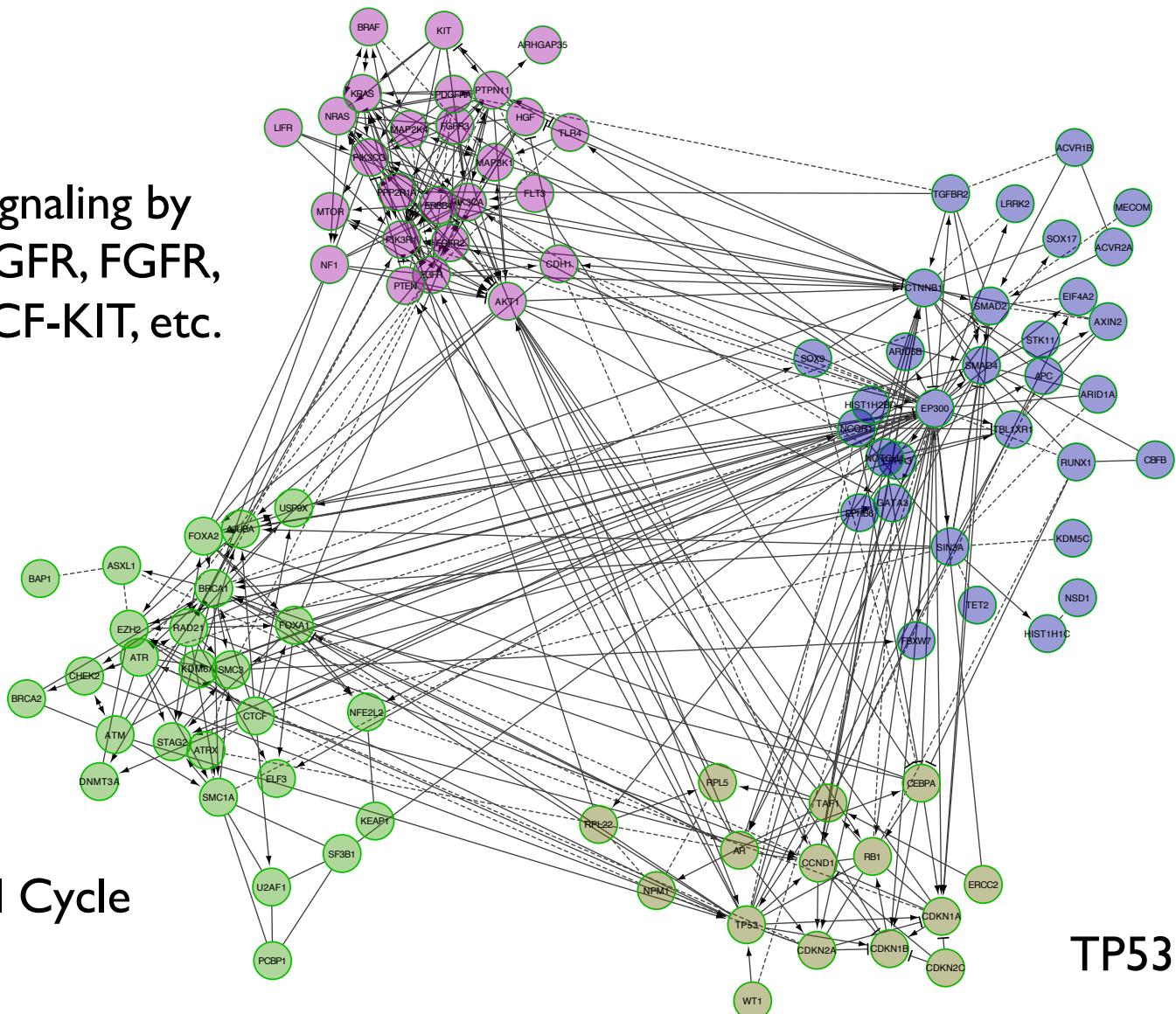
NYU Langone  
MEDICAL CENTER



OICR  
Ontario Institute  
for Cancer Research

# 127 Cancer Driver Genes

# Signaling by EGFR, FGFR, SCF-KIT, etc.

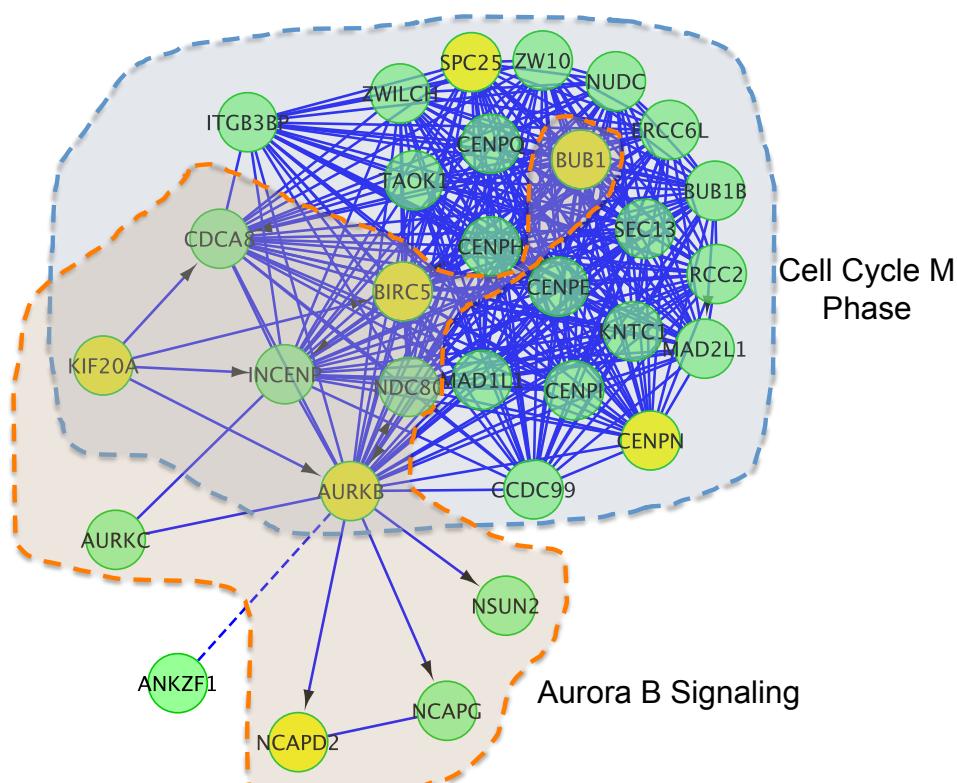


# Signaling by Notch, Wnt, TGF-beta, SMAD2/3, etc.

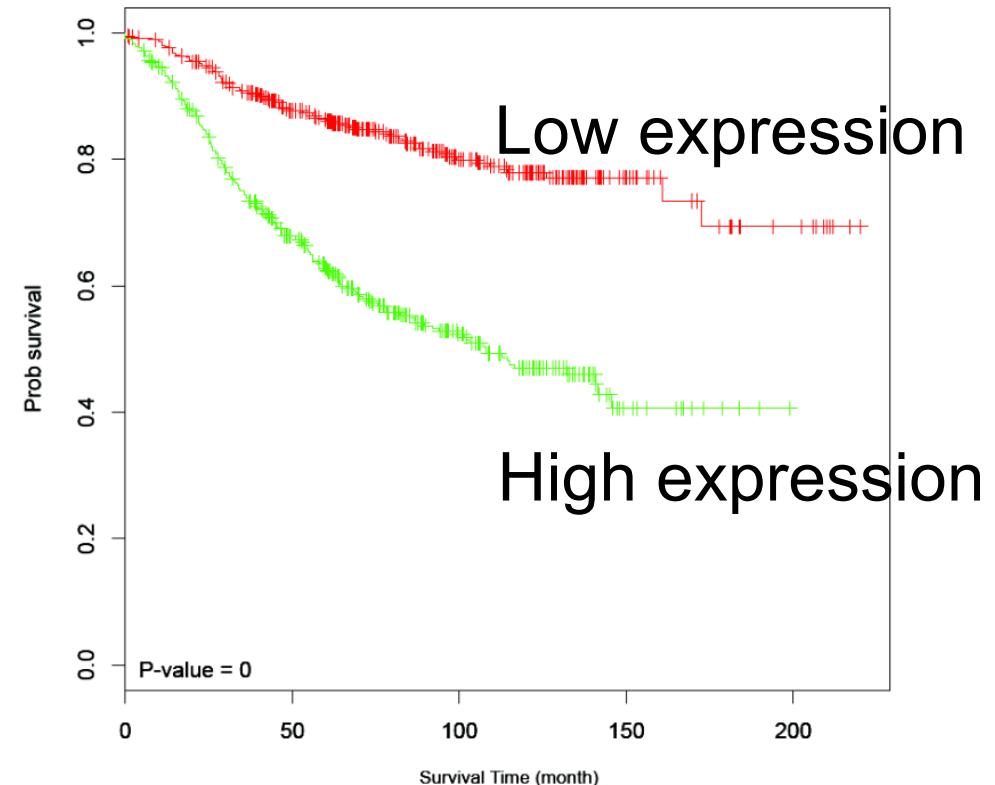
# Cell Cycle

# TP53 Pathway

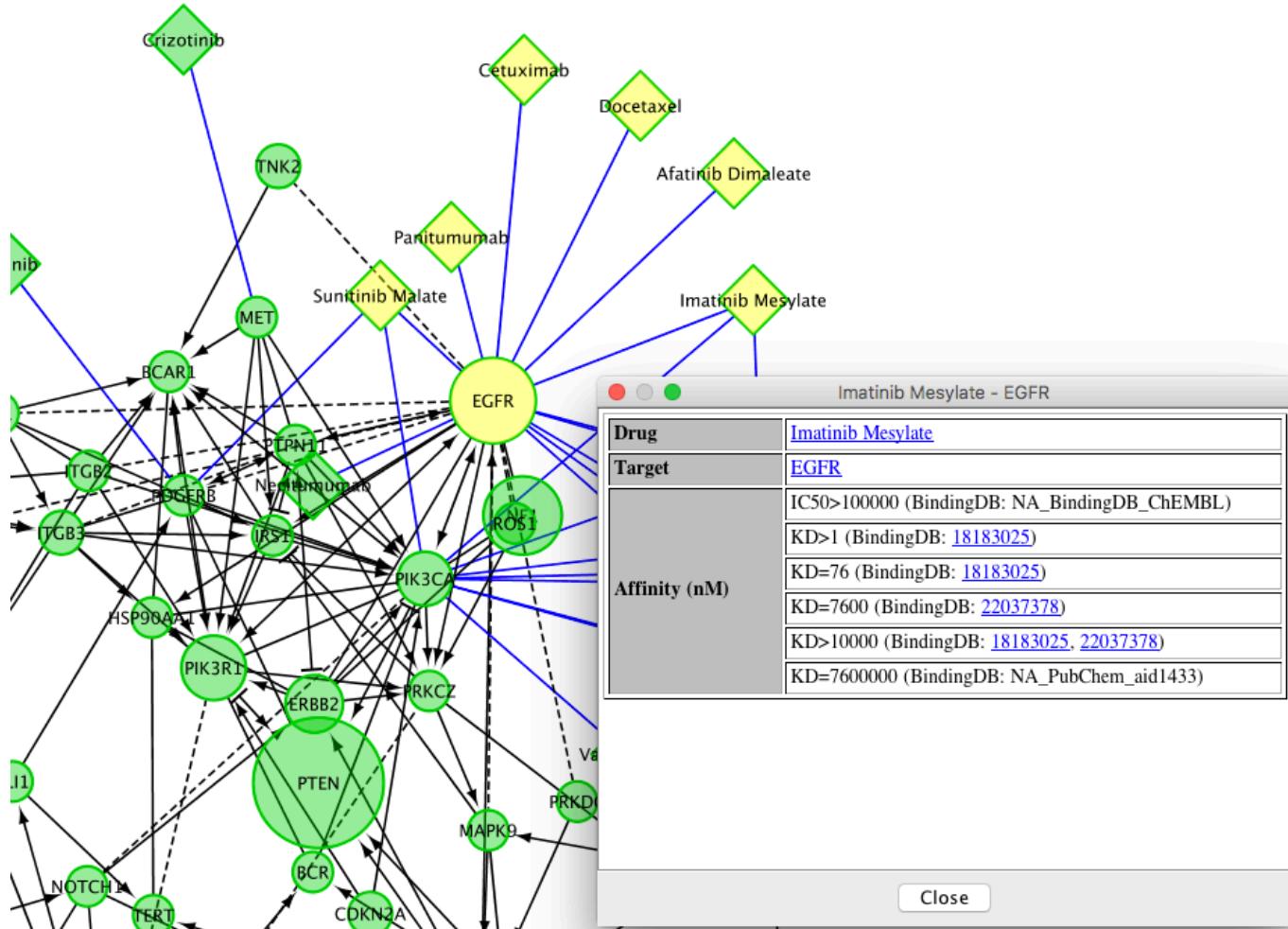
# Module-Based Prognostic Biomarker in ER+ Breast Cancer



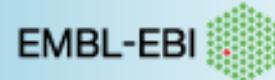
Measure levels of expression  
of the genes in this network  
module



# Visualize Cancer Targetome in the Reactome FI Network



TCGA GBM Mutation Profile



### 3) Pathway-Based Modeling

- Apply list of altered {genes, proteins, RNAs} to biological pathways.
- Preserve detailed biological relationships.
- Attempt to integrate multiple molecular alterations together to yield lists of altered pathway activities.
- Pathway modeling shades into Systems Biology

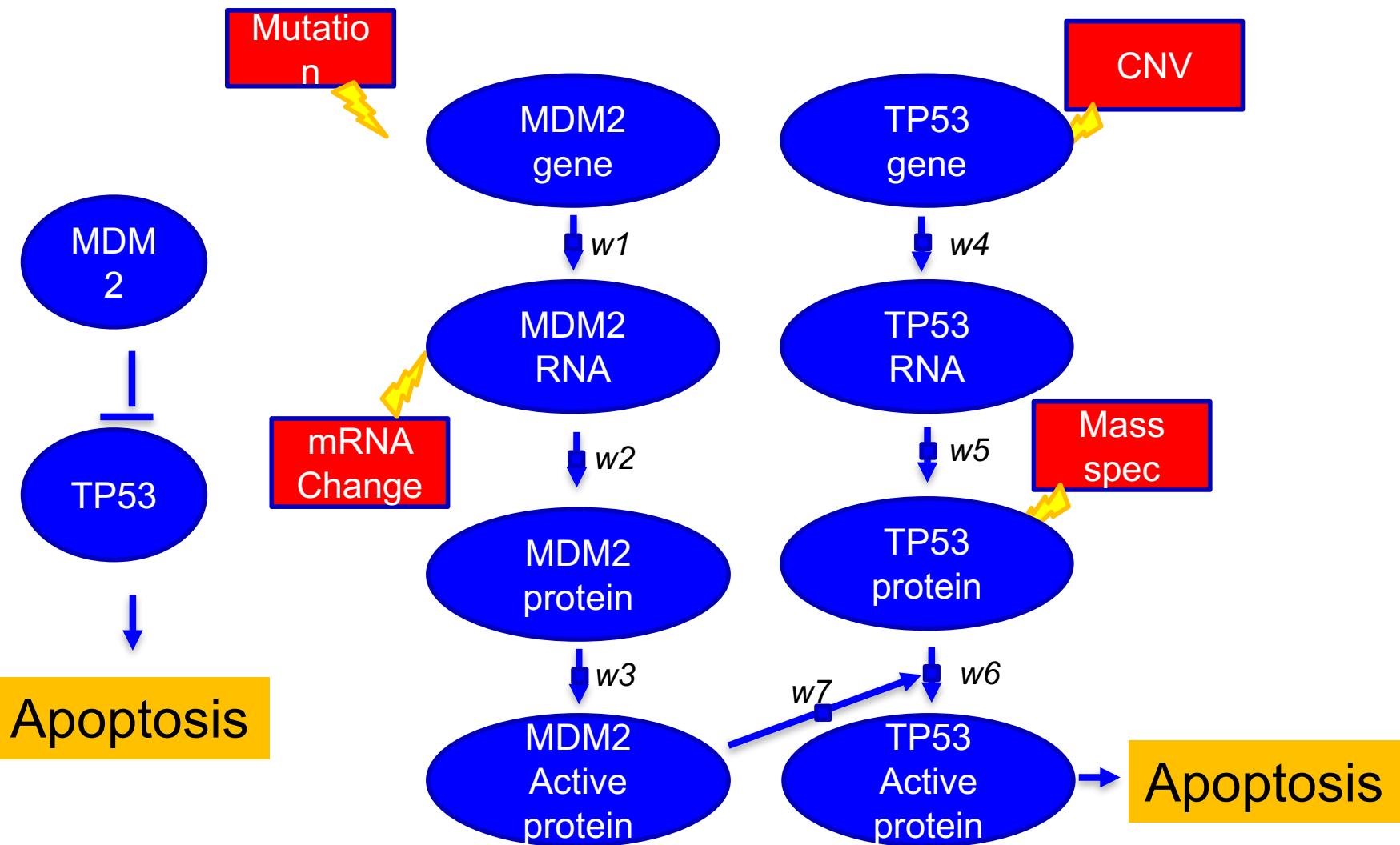
# Types of Pathway-Based Modeling

- Partial differential equations/boolean models, e.g. CellNetAnalyzer
  - Mostly suited for biochemical systems (metabolomics)
- Network flow models, e.g. 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.

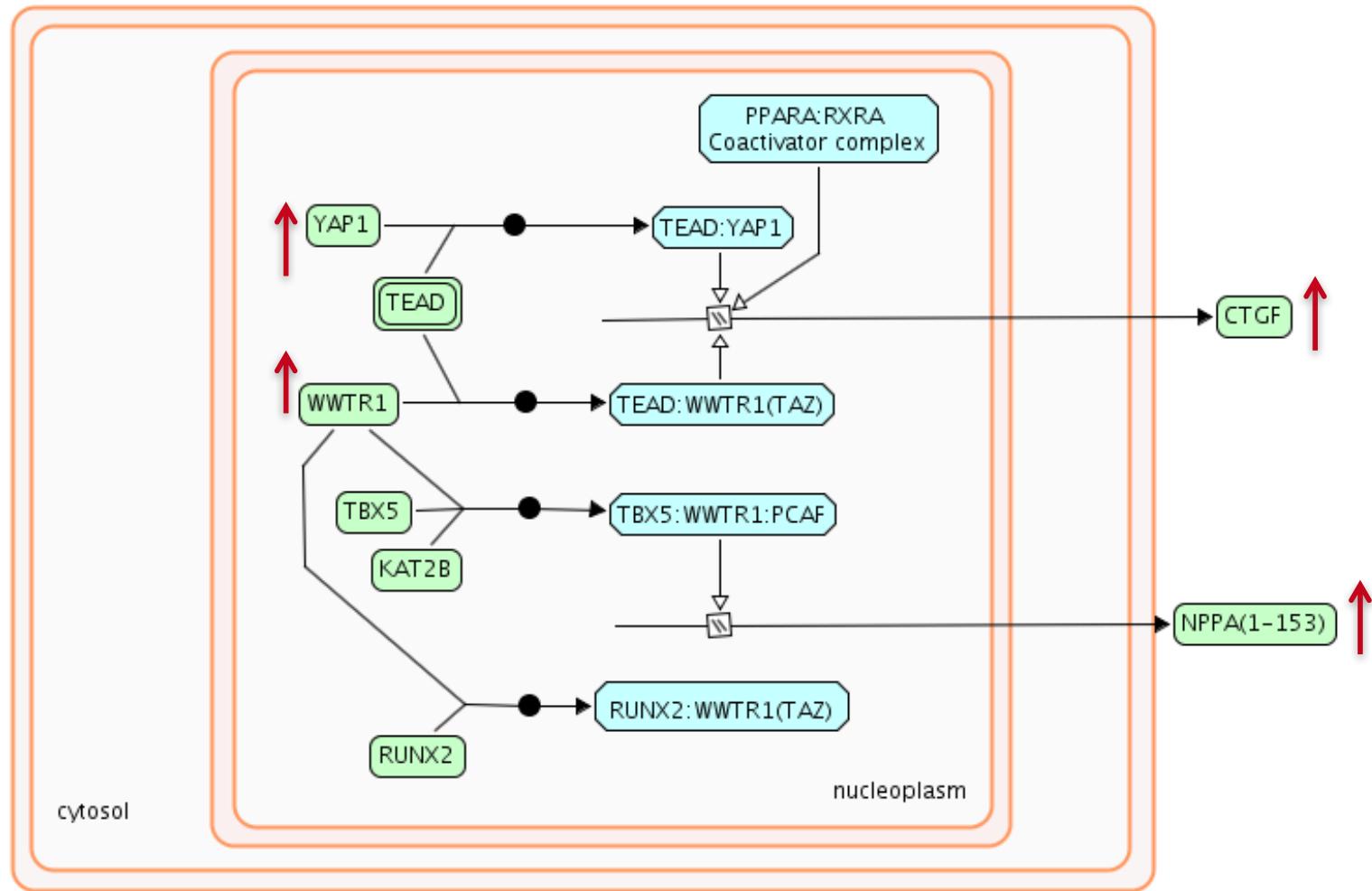
# Probabilistic Graphical Model (PGM) based Pathway Analysis

- 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?

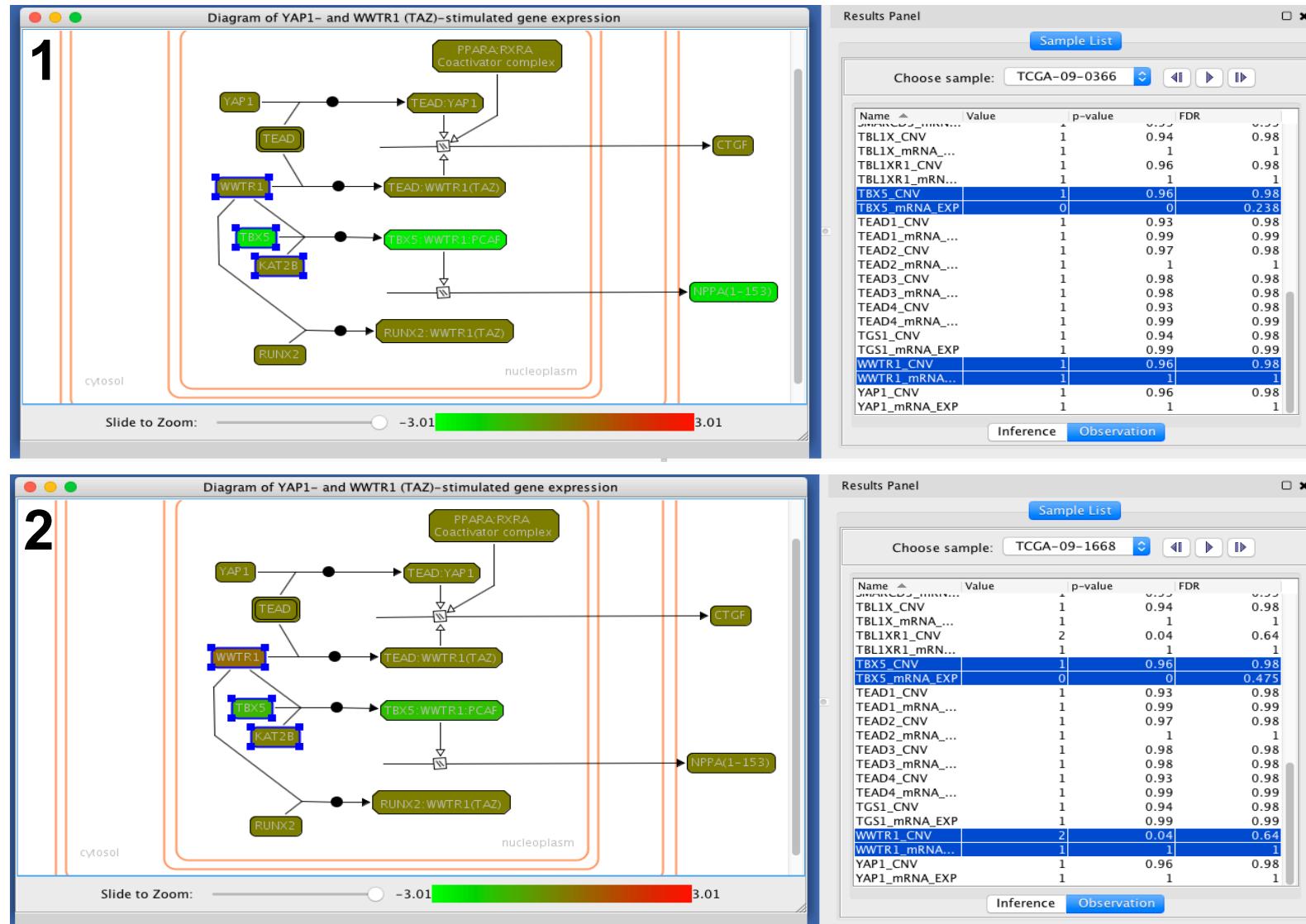
# PARADIGM



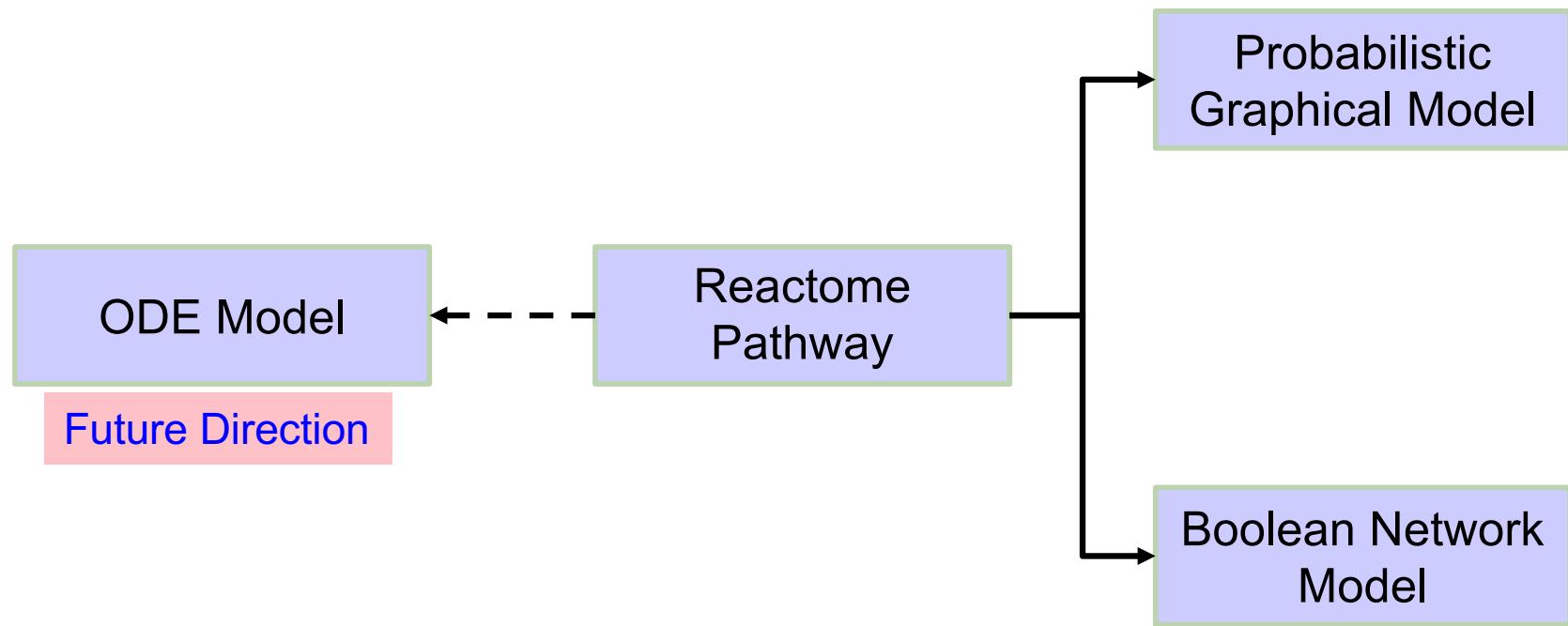
# Probabilistic Graphical Models (PGMs) for Reactome Pathways



# PGM-based Single Patient Pathway View



# Building Quantitative Models for Reactome Pathways



# Pathway/Network Database URLs

- BioGRID
  - <http://www.thebiogrid.org>
- IntAct
  - <http://www.ebi.ac.uk/intact/>
- KEGG
  - <http://www.genome.jp/kegg>
- MINT
  - <http://mint.bio.uniroma2.it>
- Reactome
  - <http://www.reactome.org>
- Pathway Commons
  - <http://www.pathwaycommons.org>
- Wiki Pathways
  - <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
  - <http://www.ebi.ac.uk/research/saez-rodriguez/software>
- NetPhorest/NetworKIN
  - <http://netphorest.info>, <http://networkin.info>
- ARACNe
  - <http://wiki.c2b2.columbia.edu/califanolab/index.php/Software/ARACNE>
- PARADIGM
  - <http://paradigm.five3genomics.com/>