

# Computational analysis of pathways and interaction networks

Gary Bader  
 Terrence Donnelly Center for Cellular and Biomolecular Research  
 (DCCBR)  
 BBDMR, SLRI  
 University of Toronto  
 JTB2010H

<http://baderlab.org>




---

---

---

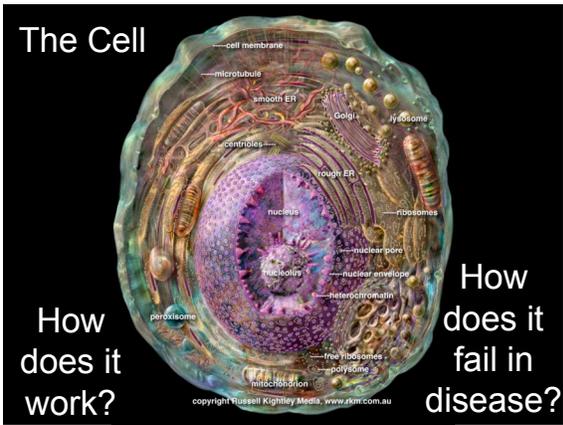
---

---

---

---

---




---

---

---

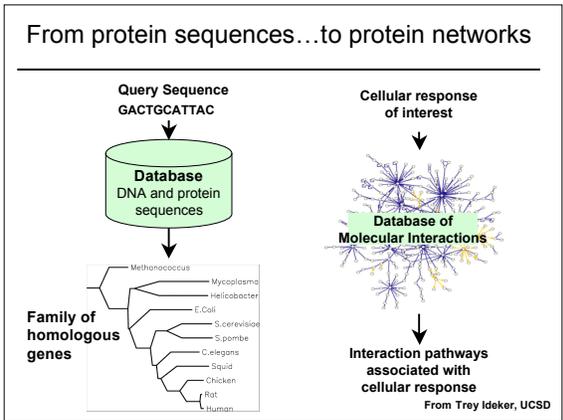
---

---

---

---

---




---

---

---

---

---

---

---

---





## Biological Pathways?

- Pathways are biological processes
- But, not really pathways -> networks
- Metabolic, signaling, regulatory and genetic
- Define gene function at many different levels
  - Human constructs to organize the cell
- A set or series of interactions, often forming a network, which biologists have found useful to group together for organizational, historic, biophysical or other reasons.

Note: generally out of cell context



---

---

---

---

---

---

---

---

## Pathways

- However, there is no *precise* biological definition of a pathway
  - You can't see a pathway 'object' under a microscope
- Start point is important or easily accessible stimulus (e.g. EGF hormone, drug)
- End point is a chosen readout (e.g. reporter gene expression, protein phosphorylation, production of a metabolite)

---

---

---

---

---

---

---

---

## What's Missing?

- Pathway/network abstraction is not dynamic
  - Difficult to represent a calcium wave or a feedback loop
- More detailed mathematical representations exist that handle these e.g.
  - Stoichiometric modeling
    - flux balance analysis, extreme pathways
  - Kinetic modeling (CyberCell, E-cell, ...)
    - Need to accumulate or estimate comprehensive kinetic information
- Level of detail
- Context

---

---

---

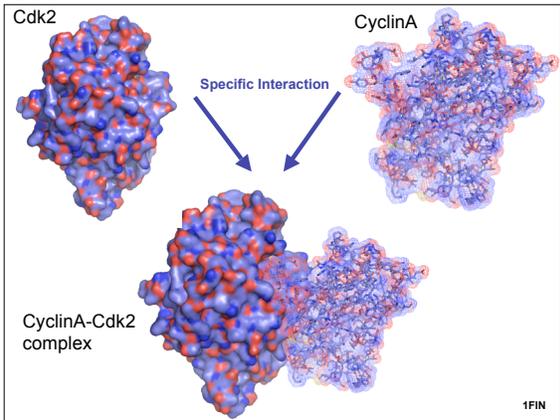
---

---

---

---

---




---

---

---

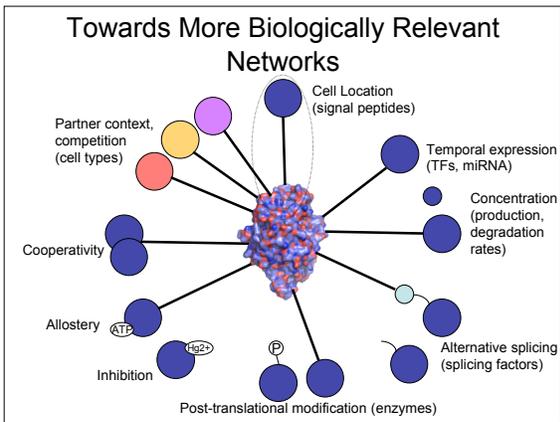
---

---

---

---

---




---

---

---

---

---

---

---

---

### Pathway Information

- **Databases**
  - Fully electronic
  - Easily computer readable
- **Literature**
  - Increasingly electronic
  - Human readable
- **Biologist's brains**
  - Richest data source
  - Limited bandwidth access
- **Experiments**
  - Basis for models

---

---

---

---

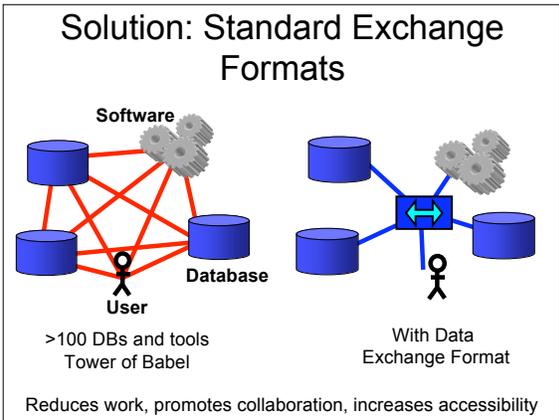
---

---

---

---






---

---

---

---

---

---

---

---

- ### Interaction and Pathway Data Exchange Formats
- PSI-MI <http://psidev.sourceforge.net>
    - Molecular interactions - protein-protein interaction focus
    - Peer reviewed, HUPPO community standard
  - BioPAX <http://www.biopax.org>
    - Biological pathways
    - Community ontology in OWL, Protégé
  - SBML <http://www.sbml.org>
    - Widely adopted for representing mathematical models of biological processes e.g. biochemical reaction networks
  - CellML <http://www.cellml.org>
    - Math models of biological processes

---

---

---

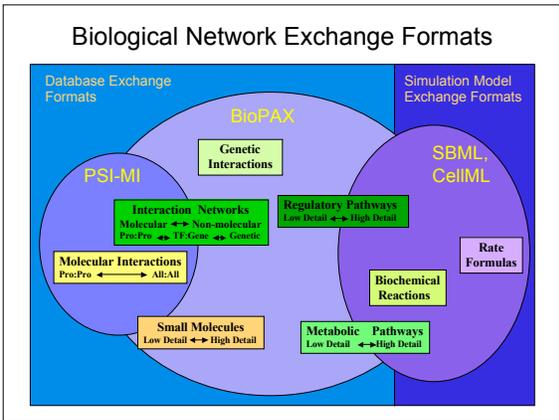
---

---

---

---

---




---

---

---

---

---

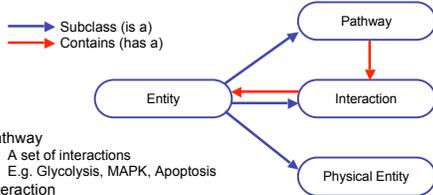
---

---

---



## BioPAX Structure



- Pathway
  - A set of interactions
  - E.g. Glycolysis, MAPK, Apoptosis
- Interaction
  - A basic relationship between a set of entities
  - E.g. Reaction, Molecular Association, Catalysis
- Physical Entity
  - A building block of simple interactions
  - E.g. Small molecule, Protein, DNA, RNA

---

---

---

---

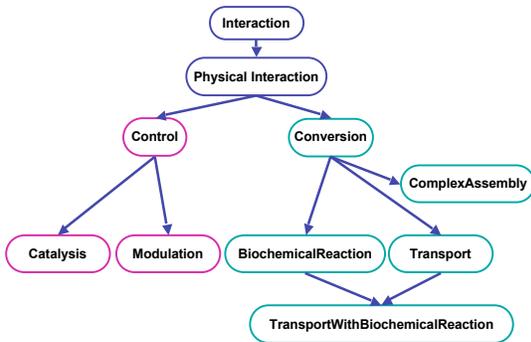
---

---

---

---

## BioPAX: Interactions




---

---

---

---

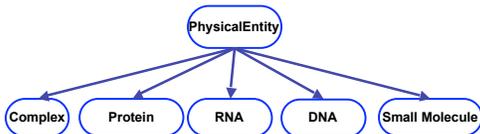
---

---

---

---

## BioPAX: Physical Entities




---

---

---

---

---

---

---

---







# The Cancer Cell Map

- EGF, TGFR, AR, Delta-notch, A6B4 Integrin, Id, Kit, TNF-alpha, Wnt, Hedgehog (10 pathways)

<http://cancer.cellmap.org>

- Details on interaction, reactions, post-translational modifications from membrane to nucleus
- Derived from original articles
- Reviewed by MSKCC experts in Massague, Benezra, Besmer, Gerald, Giancotti labs

---

---

---

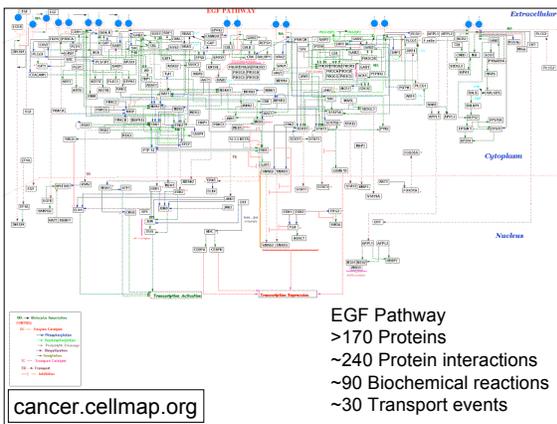
---

---

---

---

---




---

---

---

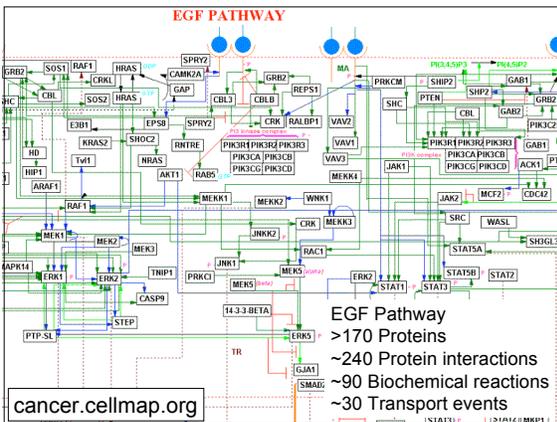
---

---

---

---

---




---

---

---

---

---

---

---

---

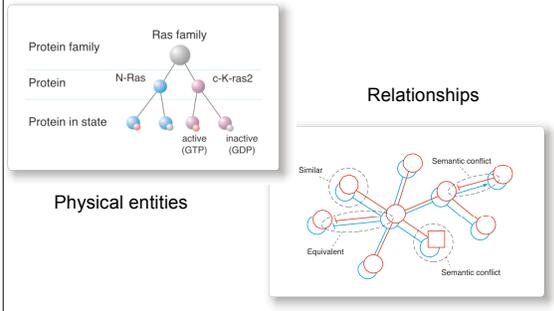






## Towards an Integrated Cell Map

- Semantic pathway integration is difficult




---

---

---

---

---

---

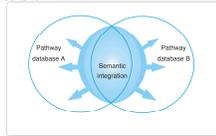
---

---

## Practical Semantic Integration

- Minimize errors
  - Integrate only where possible with high accuracy
  - Detect and flag conflicts, errors for users, no revision
  - Promote best-practices to minimize future errors
  - Interaction confidence algorithms
  - Validation software
  - Allow users to filter and select trusted sources
- Converge to standard representation
  - Community process

Doable: hundreds of curators globally in >200 databases (GDP) - make it more efficient




---

---

---

---

---

---

---

---

## Practical uses of pathway information

- Use as a reference
  - Summary of what is known about a cellular process or gene
  - Visual communication
- Analyze molecular profiles
  - After ranking and clustering
  - Summary of ranked data or clusters (1000 genes, now what?)
- Visualize molecular profiles in pathway context
- Find active pathways in sets of genes defined by ranking or clustering
- Find protein complexes in protein interaction networks

---

---

---

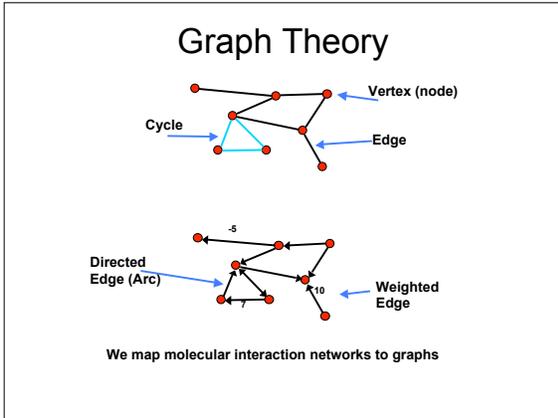
---

---

---

---

---




---

---

---

---

---

---

---

---

- ### Mapping Biology to a Graph
- A simple mapping
    - one compound/node, one interaction/edge
  - A more realistic mapping
    - Cell localization, cell cycle, cell type, taxonomy
    - Only represent physiologically relevant interaction networks
  - Edges can represent other relationships
  - **Critical:** understand the mapping for network analysis

---

---

---

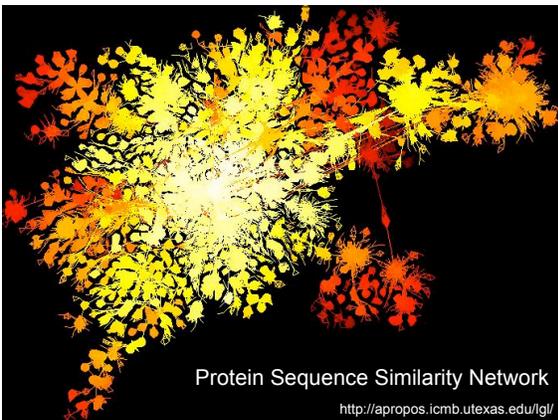
---

---

---

---

---




---

---

---

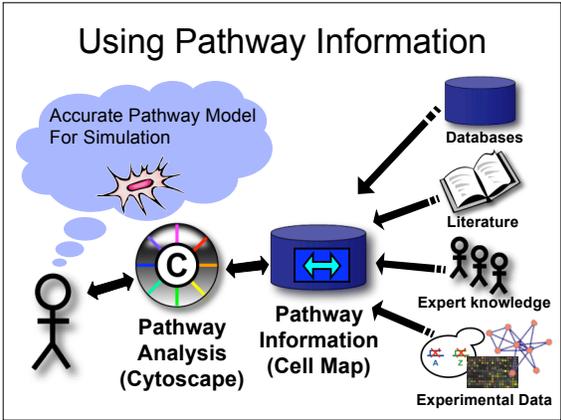
---

---

---

---

---




---

---

---

---

---

---

---

---

**Cytoscape**  
<http://cytoscape.org>

**UCSF**

**Agilent Technologies**

---

---

---

---

---

---

---

---

### Cytoscape - Network Visualization and Analysis

<http://cytoscape.org>

- Freely-available (open-source, java) software
- Visualizing biological networks (e.g. molecular interaction networks)
- Analyzing networks with gene expression profiles and other cell state data

Other software: Osprey, BioLayout, VisANT, Navigator, PIMWalker, ProViz

---

---

---

---

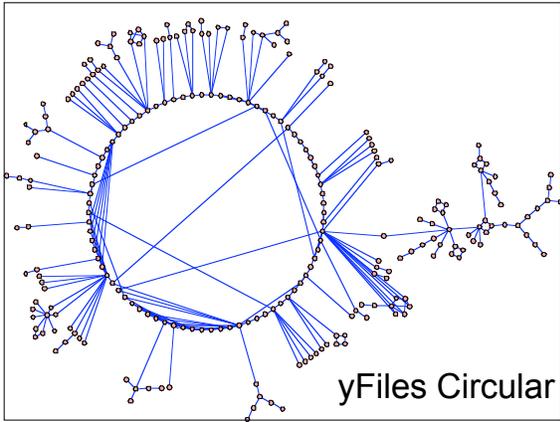
---

---

---

---






---

---

---

---

---

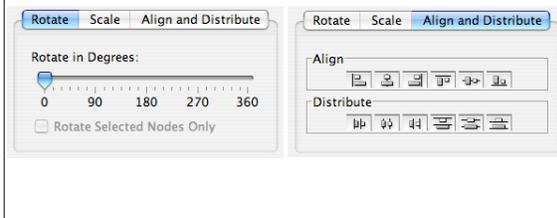
---

---

---

## Network Layout

- 15 algorithms available through plugins
- Demo: Move, zoom/pan, rotate, scale, align




---

---

---

---

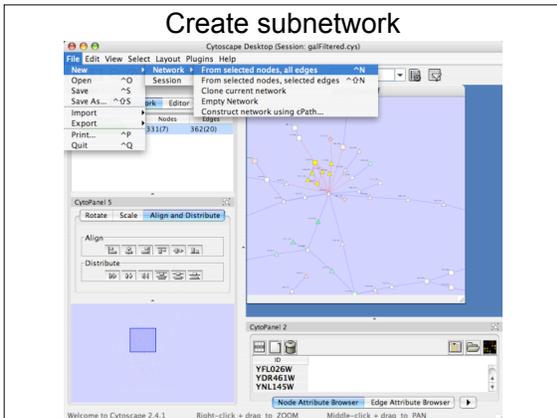
---

---

---

---

## Create subnetwork




---

---

---

---

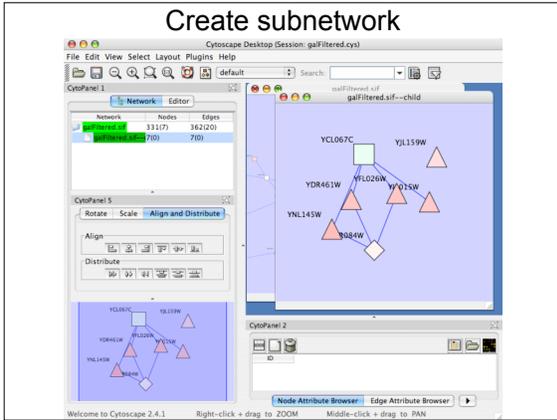
---

---

---

---

## Create subnetwork



---

---

---

---

---

---

---

---

## Visual Style

- Customized views of experimental data in a network context
- Network has node and edge attributes
  - E.g. expression data, GO function, interaction type
- Mapped to visual attributes
  - E.g. node/edge size, shape, colour...
- E.g. Visualize gene expression data as node colour gradient on the network

---

---

---

---

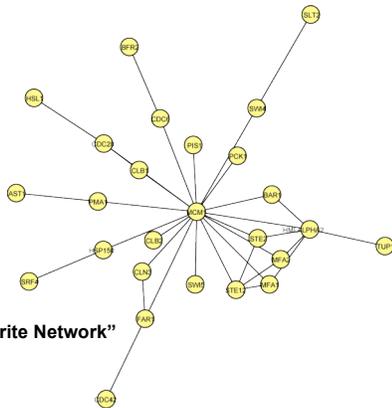
---

---

---

---

## Visual Style



Load "Your Favorite Network"

---

---

---

---

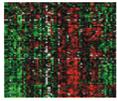
---

---

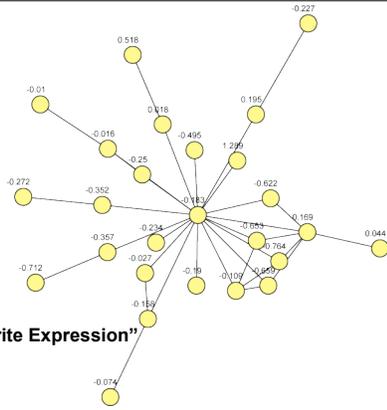
---

---

## Visual Style



Load "Your Favorite Expression" Dataset



---

---

---

---

---

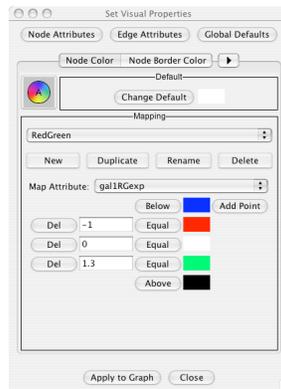
---

---

---

## Visual Style

Map expression values to node colours using a continuous mapper



---

---

---

---

---

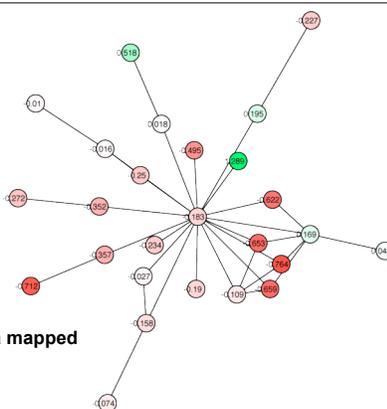
---

---

---

## Visual Style

Expression data mapped to node colours



---

---

---

---

---

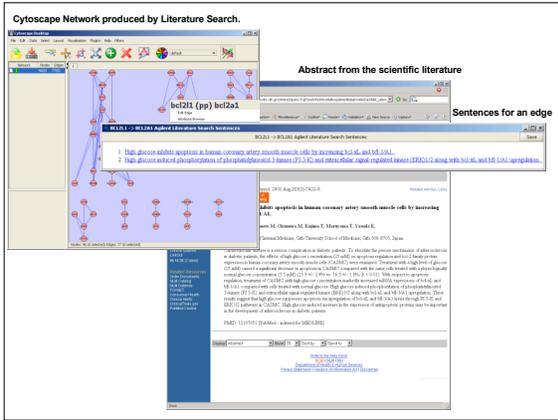
---

---

---








---

---

---

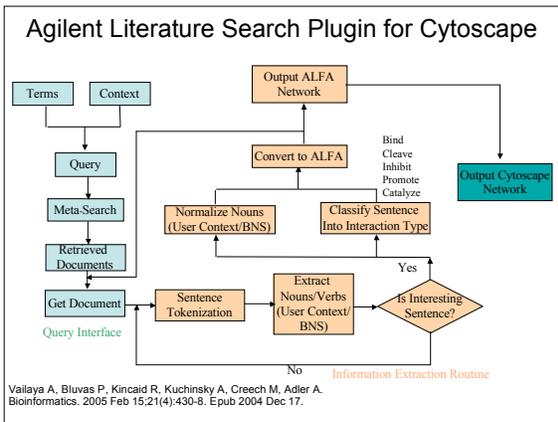
---

---

---

---

---




---

---

---

---

---

---

---

---

## Gene Expression/Network Integration

- Identifier (ID) mapping
  - Translation from network IDs to gene expression IDs e.g. Affymetrix probe IDs
  - Also: Unification, link out, query
  - Entrez gene IDs (genes), UniProt (proteins)
- Synergizer
  - [llama.med.harvard.edu/cgi/synergizer/translate](http://llama.med.harvard.edu/cgi/synergizer/translate)
- More ID mapping services available
  - <http://baderlab.org/IdentifierMapping>

---

---

---

---

---

---

---

---

# Gene Expression/Network Integration

## THE SYNERGIZER

The Synergizer database is a growing repository of gene and protein identifier synonym relationships. This tool facilitates the conversion of identifiers from one naming scheme (a.k.a "namespace") to another.

load sample inputs

Select species:

Select authority:

Select "FROM" namespace:

Select "TO" namespace:

File containing IDs to translate:

and/or

IDs to translate:

Output as spreadsheet:

(NB: The strings in [brackets] are representative IDs in the corresponding namespaces.)

	* <b>entrezgene</b>
Y1L062C	854748
Y1L370C	851085
Y1K013C	833856
YNR035C	855771
YBR234C	852536

1. Load as attributes in Cytoscape

2. Assign expression values to nodes using this attribute set

---

---

---

---

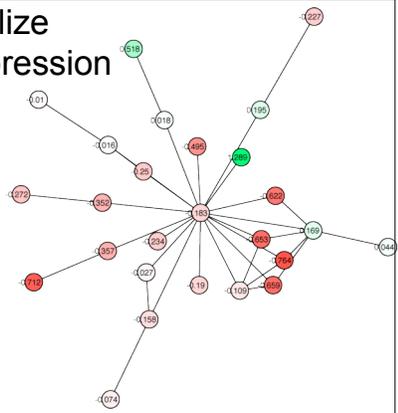
---

---

---

---

## Visualize Gene Expression




---

---

---

---

---

---

---

---

## Find Active Subnetworks

- Active modules
  - Input: network + p-values for gene expression values e.g. from GCRMA
  - Output: significantly differentially expressed subgraphs
- Method
  - Calculate z-score/node,  $Z_A$  score/subgraph, correct vs. random expression data sampling
  - Score over multiple experimental conditions
  - Simulated annealing used to find high scoring networks

Ideker T, Ozier O, Schwikowski B, Siegel AF Bioinformatics. 2002;18 Suppl 1:S233-40

---

---

---

---

---

---

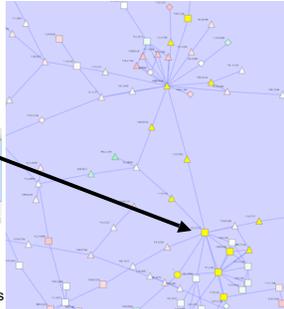
---

---

## Active Module Results

Network: yeast protein-protein and protein-dna network  
 Expression data: 3 gene knock out conditions (enzyme, TF activator, TF repressor)

Conditions vs. Pathways					
Network	Size	Score	gal1ΔCug	gal4ΔCug	gal80ΔCug
1	14	3.78			
2	26	3.584			
3	10	2.994			
4	7	2.934			
5	4	2.634			



Note: non-deterministic, multiple runs required for confidence of result robustness

Ideker T et al. Science. 2001 May 4;292(5518):929-34.

---

---

---

---

---

---

---

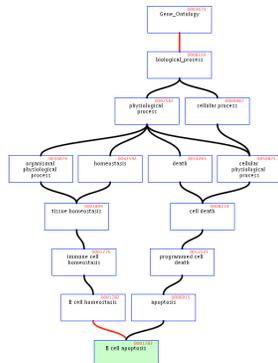
---

---

---

## The Gene Ontology (GO)

- Describes gene function
- 1. Agreed upon terms (controlled vocabulary)
  - Biological process
  - Cellular component
  - Molecular function
- 2. Genome annotation



[www.geneontology.org](http://www.geneontology.org)

---

---

---

---

---

---

---

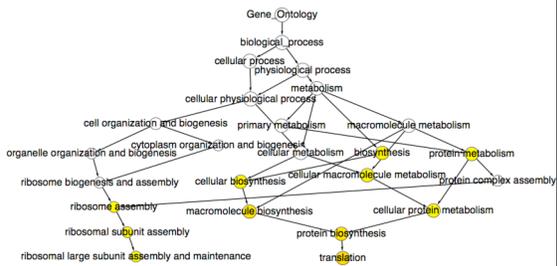
---

---

---

## BiNGO

Hypergeometric p-value  
 Multiple testing correction  
 (Benjamini-Hochberg FDR)



Caveats: Gene identifiers must match;  
 low GO term coverage, GO bias

Maere, S., Heymans, K. and Kuiper, M  
 Bioinformatics 21, 3448-3449, 2005

---

---

---

---

---

---

---

---

---

---

## NetMatch

- Query a network for topological matches
- Input: query and target networks, optional node/edge labels
- Output: Topological query matches as subgraphs of target network
- Supports: subgraph matching, node/edge labels, label wildcards, approximate paths
- <http://alpha.dmi.unict.it/~ctnyu/netmatch.html>

Ferro A, Giugno R, Pigola G, Pulvirenti A, Skripin D, Bader GD, Shasha D  
 Bioinformatics 2007 Feb 3  
 Extends state space representation based search from Cordella et al. IEEE  
 Transactions on Pattern Analysis and Machine Intelligence, 2004, 26, 10, 1367–1372

---

---

---

---

---

---

---

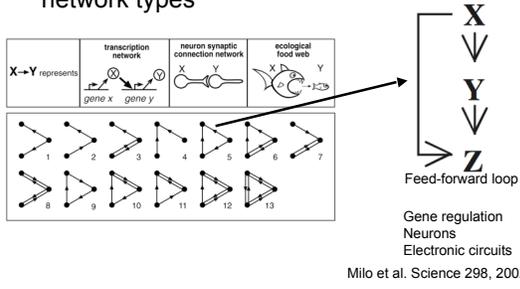
---

---

---

## Find Feed-Forward Motifs

- Graph motifs over-represented in many network types




---

---

---

---

---

---

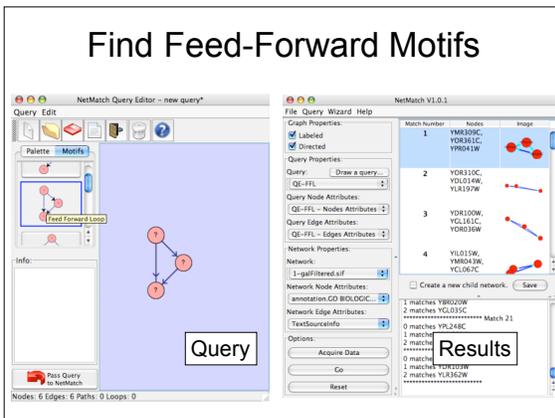
---

---

---

---

## Find Feed-Forward Motifs




---

---

---

---

---

---

---

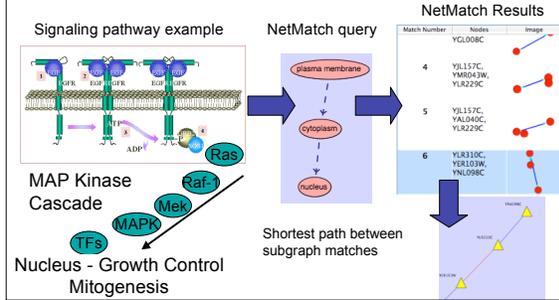
---

---

---

## Find Signaling Pathways

- Potential signaling pathways from plasma membrane to nucleus via cytoplasm




---

---

---

---

---

---

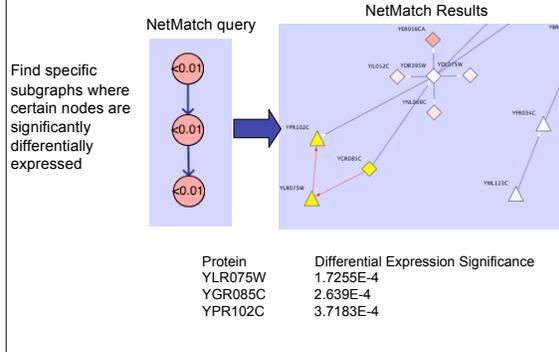
---

---

---

---

## Find Expressed Motifs




---

---

---

---

---

---

---

---

---

---

## Graph Clustering - MCODE Plugin

- Clusters in a protein-protein interaction network have been shown to represent protein complexes and parts of pathways
- Clusters in a protein similarity network represent protein families
- Network clustering is available through the MCODE Cytoscape plugin

---

---

---

---

---

---

---

---

---

---



## Other Pathway Analyses

- Find known pathways in new species using Pathblast ([www.pathblast.org](http://www.pathblast.org))
- Find active pathways from molecular profiles (e.g. active modules, activity centers, GOALIE)
- Molecular interaction and pathway prediction from genome sequence
- Pathway simulation to predict drugs and drug combinations that activate or inhibit specific biological processes

---

---

---

---

---

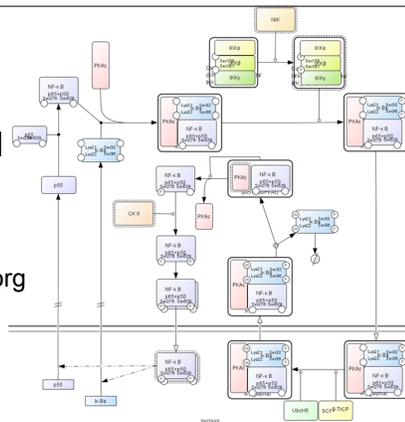
---

---

---

## Systems Biology Graphical Notation

<http://sbgn.org>  
In progress




---

---

---

---

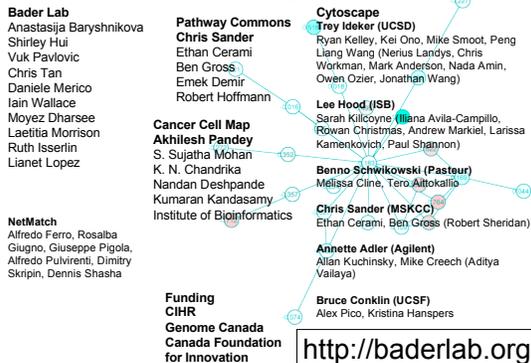
---

---

---

---

## Acknowledgements




---

---

---

---

---

---

---

---