

Integrative analysis of interaction networks

Gary Bader <http://www.baderlab.org>
JTB2010 - Nov.23.2009



Donnelly Centre
for Cellular + Biomolecular Research



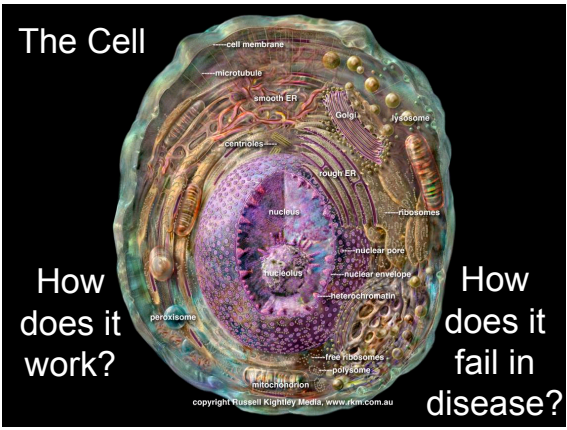
UNIVERSITY OF
TORONTO

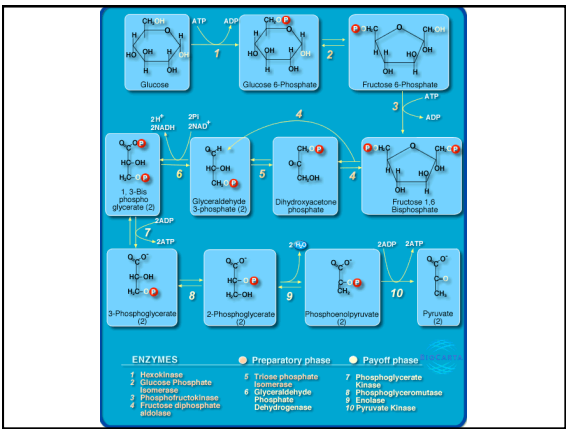


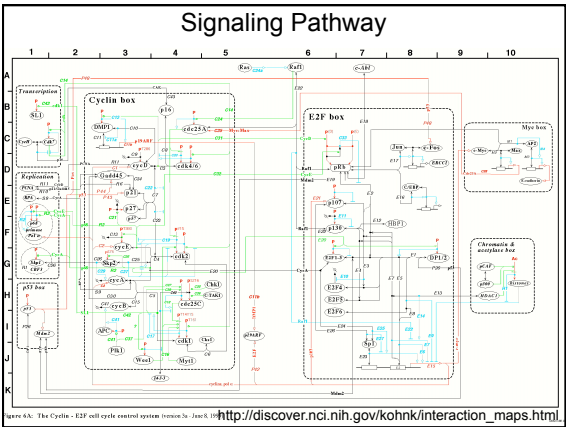
Outline

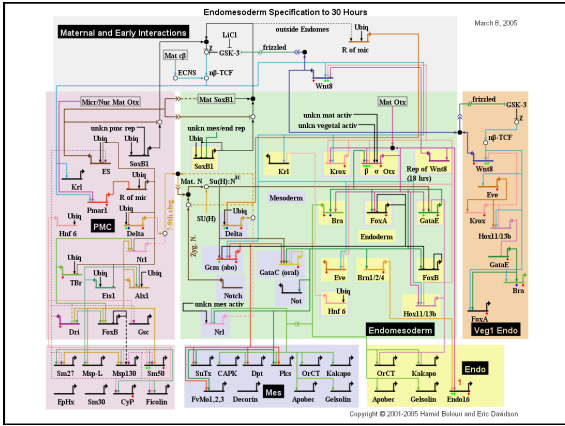
- Data integration using networks
- Network analysis
- Network data
- Network visualization and analysis
- Analyzing molecular profiles

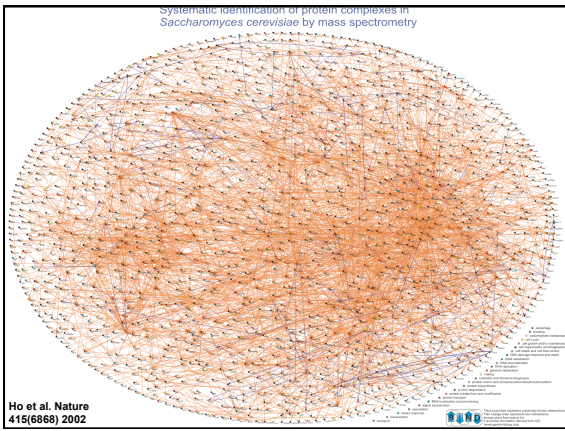
Data integration using networks

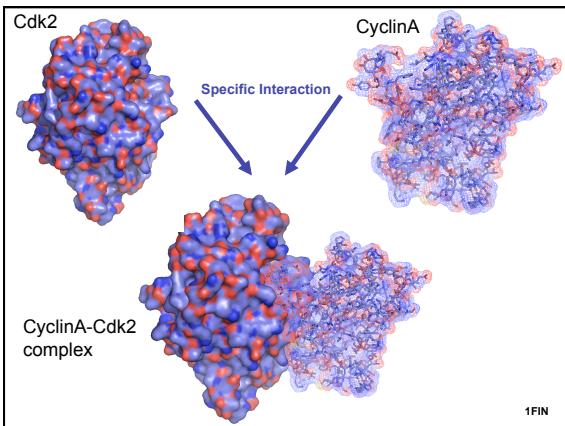


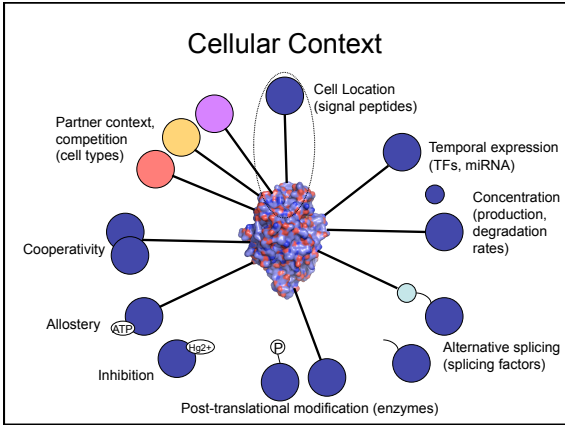


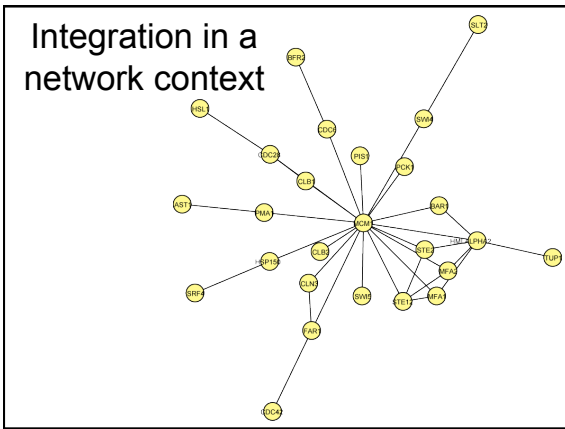


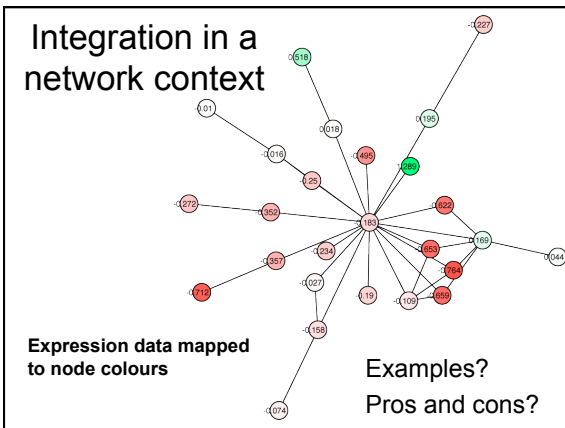












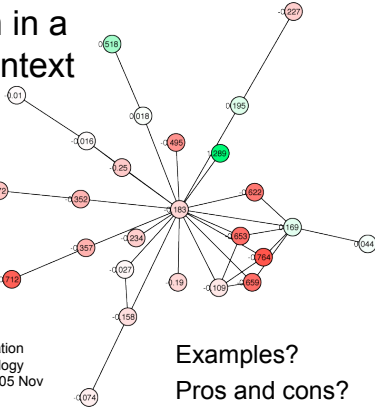
Expression data mapped to node colours

Integration in a network context

Advantages:
 -Interpretable
 -Broader coverage
 -Error reduction

Challenges:
 -Must carefully match data sets to avoid errors e.g. different interaction experiments
 -Consider data set bias
 -Consider binary vs. discrete vs. continuous

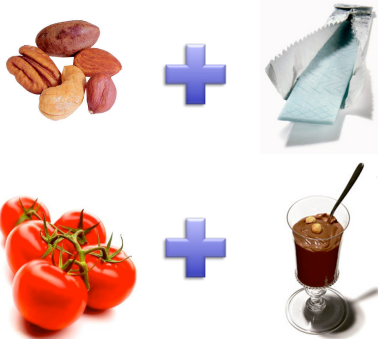
Hwang D et al. A data integration methodology for systems biology
 Proc Natl Acad Sci U S A. 2005 Nov 29;102(48):17296-301



Examples?
 Pros and cons?



Data Integration



Network Analysis

Why Network Analysis?

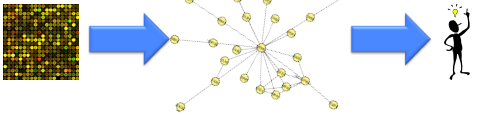
Intuitive to Biologists

- Provide a biological context for results
- More efficient than searching databases gene-by-gene
- Intuitive display for sharing data

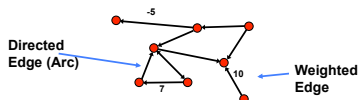
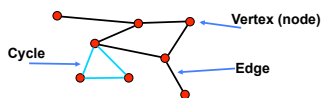
Computationally Query to Answer Specific Questions

- Visualize multiple data types on a network
- Cluster, Find active pathways, Compare, Search

Eureka! New pathway gene!



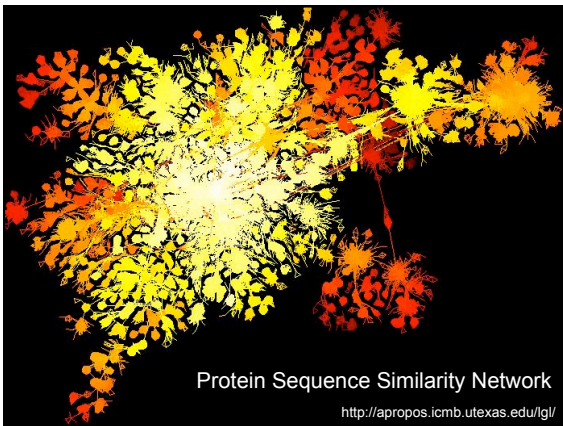
Graph Theory



We map molecular interaction networks to graphs

Mapping Biology to a Network

- 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

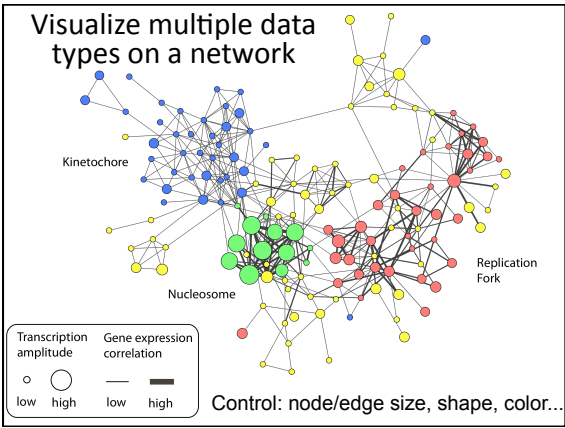


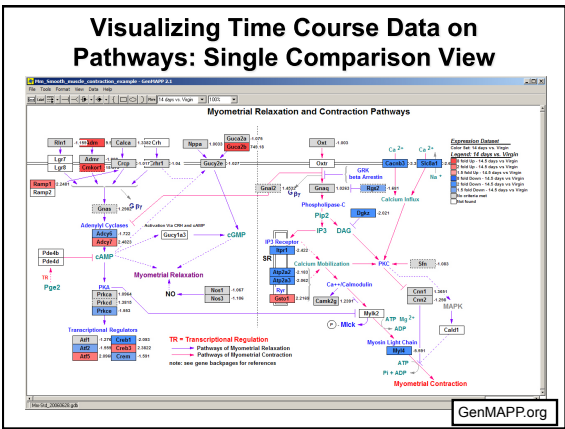
Six Degrees of Separation

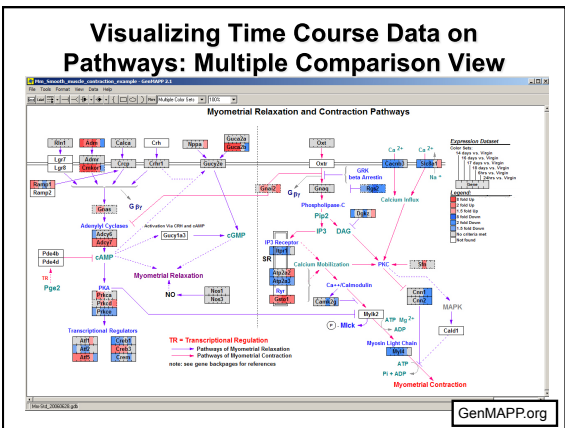
- Everyone in the world is connected by at most six links
- Which path should we take?
- Shortest path by breadth first search
 - If two nodes are connected, will find the shortest path between them
- Are two proteins connected? If so, how?
- Biologically relevant?



<http://www.time.com/time/techtme/200406/community.html>

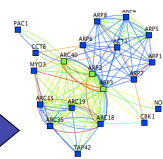






Predicting Gene Function

arp2
arp3
arc40



• STRING
– <http://string.embl.de/>

• bioPIXIE
– <http://pixie.princeton.edu/pixie/>

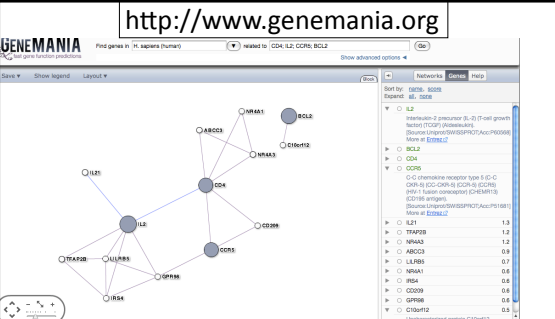
• GeneMania
– <http://www.genemania.org>

Fraser AG, Marcotte EM - A probabilistic view of gene function - Nat Genet. 2004 Jun;36(6):559-64

Top-Scoring Genes

AFG1B	0.09026
AFG1P	0.08677
AFG2B	0.08514
AFG1B	0.07793
AFG1P	0.03239
AFG1B	0.02344
AFG1P	0.02293
AFG1B	0.02031
TAK1L2	0.02017
ACT1	0.01854
AFG1A	0.01841
AFG1A	0.01792
NGC2	0.01676
PAC1	0.01563
AFG1T	0.01561
MYO3	0.01551

<http://www.genemania.org>



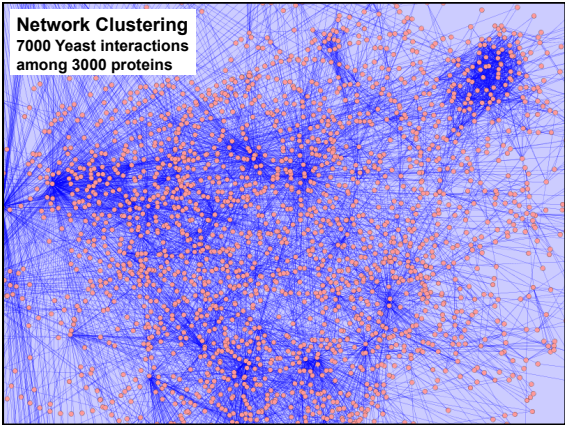
Gene Function Prediction

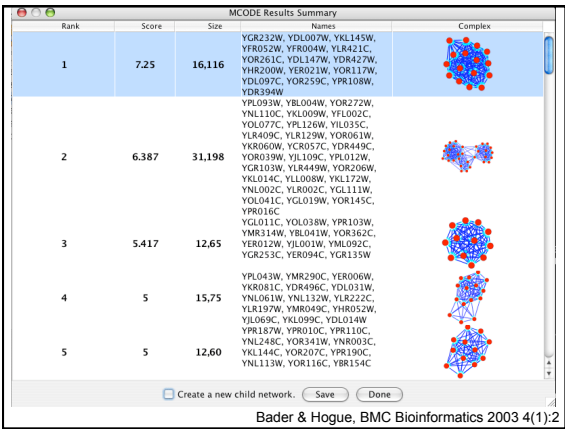
- Guilt-by-association principle
- Biological networks are combined intelligently to optimize prediction accuracy
- Algorithm is more fast and accurate than its peers

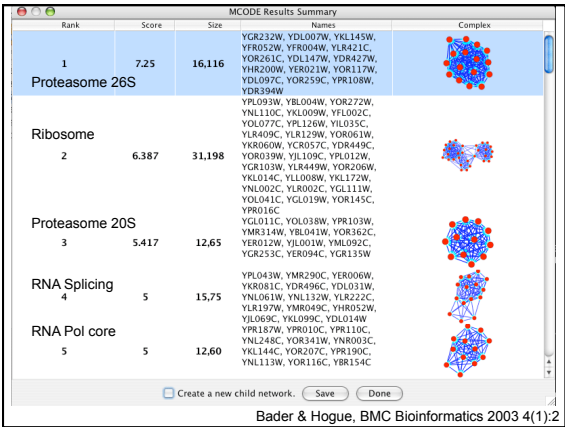
Quaid Morris (CCBR)
Rashad Badrawi, Ovi Comes, Sylvia Donaldson, Christian Lopes, Farzana Kazi, Jason Montojo, Harold Rodriguez, Khalid Zuberi

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

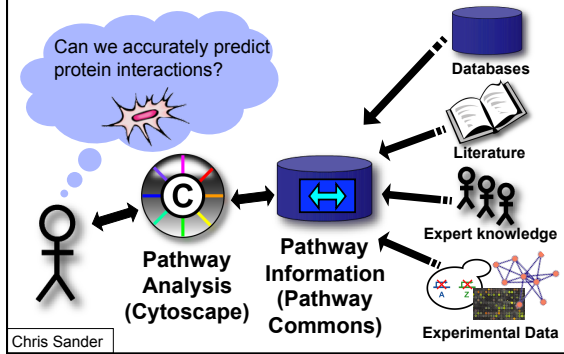






Network Data

Cell map exploration and analysis



http://pathguide.org

Pathguide the pathway resource list

Home | PubPALS | EBID | LocusMap

>300 Pathway Databases!

Complete Listing of All Pathguide Resources

Pathguide contains information about 322 biological pathway resources. Click or link to go to the resource home page or Details for a description page. Databases full and free and those supporting BiOPAX, CellML, PISA or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please send us an e-mail.

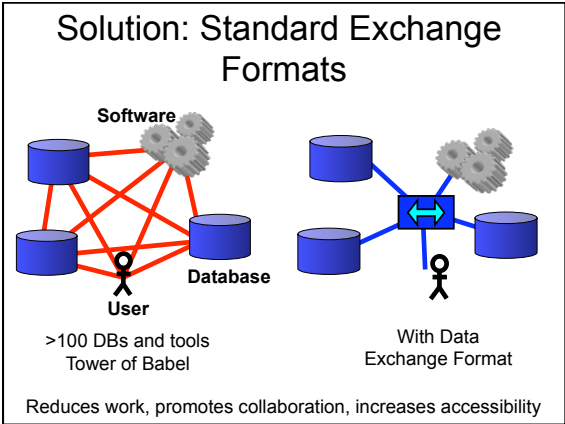
Get the State
Detailed Pathguide resource
resources now available
Pathguide Published
Check out the new resource!

Protein-Protein Interactions

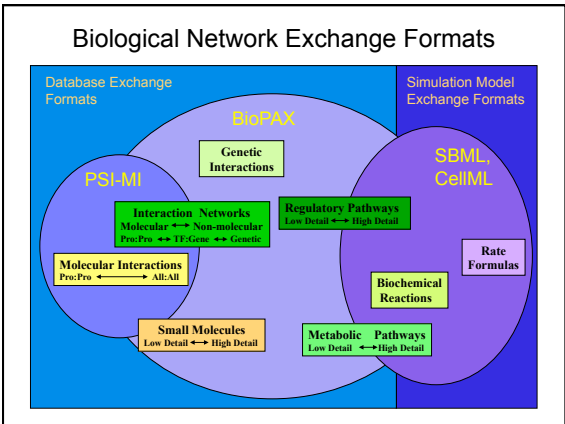
Database Name	Order	Availability	Standards
3DID - 3D interacting domains	Details	Free	
ABCDB - Arabidopsis and Bacteria ABC transporter database	Details	Free	
ACS - Alliance for Cellular Signaling Molecule Pages Database	Details	Free	
AFPase - Functional Associations of Proteins in Complete Genomes	Details	Free	
ASEdb - Alliance Screening Energetics Database	Details	Free	
ASPD - Artificial Selected Proteins/Peptides Database	Details	Free	
BID - Binding Interface Database	Details	Free	
BNID - Biomolecular Interaction Network Database	Details	Free	BiPAX
BindingDB - The Binding Database	Details	Free	
BiOGRID - General Repository for Interaction Datasets	Details	Free	BiPAX
BRITE - Biomolecular Relations in Information Transmission and Expression	Details	Free	
CA1Neuron - Pathways of the Hippocampal CA1 neuron	Details	Free	
Cancer Cell Map - The Cancer Cell Map	Details	Free	BiPAX
CSP - Cytokine Signaling Pathway Database	Details	Free	
CTDB - Calmodulin Target Database	Details	Free	
DOB - Database of Domain Interactions and Bindings	Details	Free	
DIP - Database of Interacting Proteins	Details	Free	BiPAX
DoDdb - Database of oligomer	Details	Free	
Doppel - Doppel	Details	Free	
DR - Database of ribosomal	Details	Free	
DSM - Dynamic Signaling Map	Details	Free	
FIMM - Functional Molecular Interactions	Details	Free	
FusioDB - Protein-Protein Gene P.	Details	Free	

•Varied formats, representation, coverage
•Pathway data extremely difficult to combine and use

Vuk Pavlovic



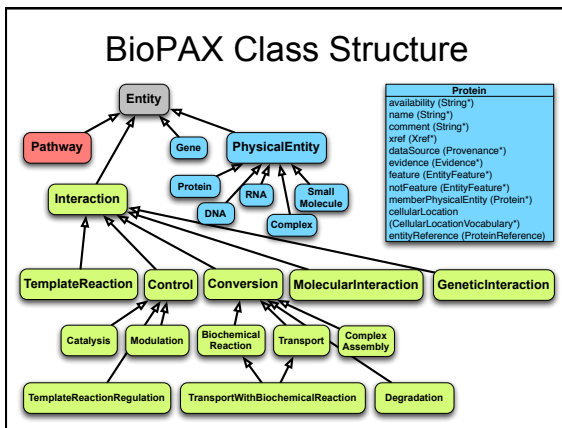
- ### Interaction and Pathway Data Exchange Formats
- PSI-MI <http://psidev.sourceforge.net>
 - Molecular interactions - protein-protein interaction focus
 - Peer reviewed, HUP0 community standard
 - BioPAX <http://www.biopax.org>
 - Biological pathways
 - Community ontology in OWL, Protégé
 - SBML <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 Pathway Language

- Represent:
 - Metabolic pathways
 - Signaling pathways
 - Protein-protein, molecular interactions
 - Gene regulatory pathways
 - Genetic interactions
- Community effort: pathway databases distribute pathway information in standard format
 - Over 100 people, database groups, standard efforts

BioPAX Class Structure

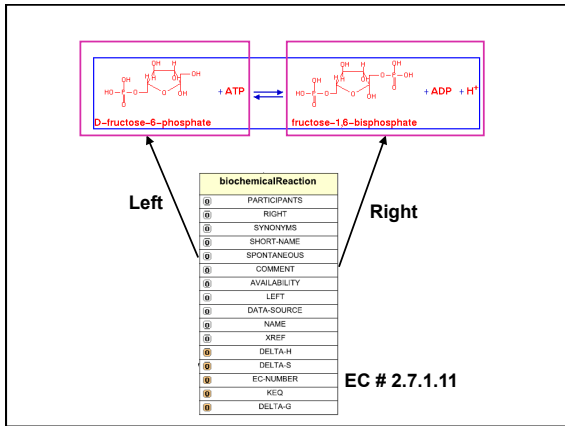


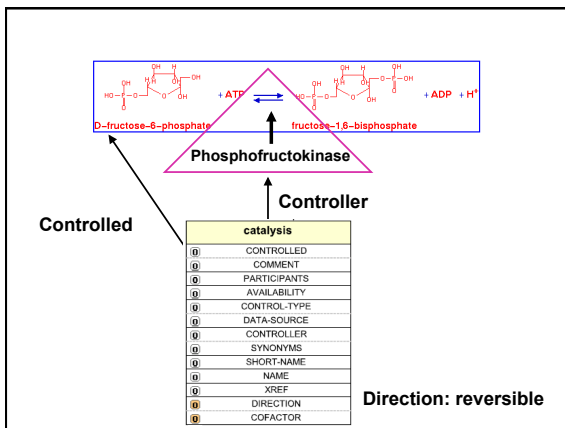


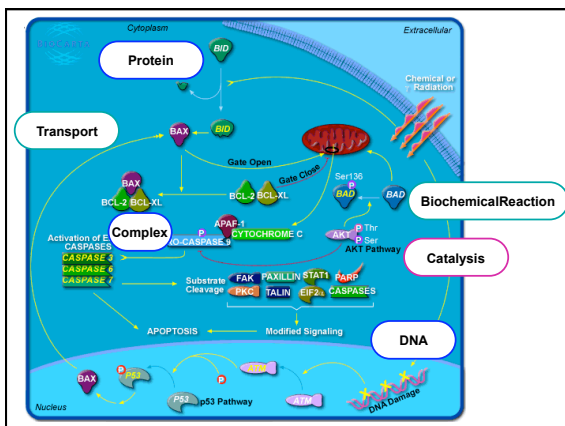
Phosphofructokinase

Biochemical Reaction
Glycolysis Pathway

Source: BioCyc.org







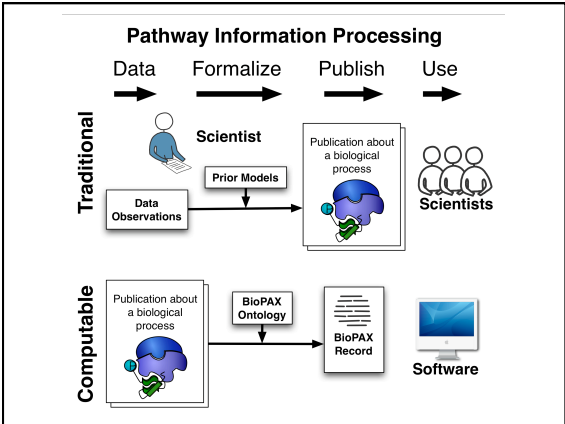
Controlled Vocabularies (CVs)

- BioPAX uses existing CVs where available via openControlledVocabulary instances
 - Cellular location: Gene Ontology (GO) component
 - PSI-MI CVs for:
 - Protein post-translational modifications
 - Interaction detection experimental methods
 - Experimental form
 - PATO phenotypic quality ontology
 - Some database providers use their own CVs
 - E.g. BioCyc evidence codes
- More at the Ontology Lookup Service
 - <http://www.ebi.ac.uk/ontology-lookup/>

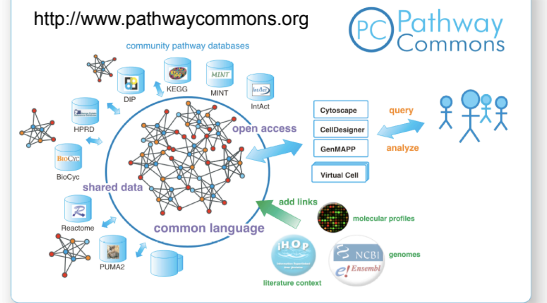
OWL (XML) Snippet

```

<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/151" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/152" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/153" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/154" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/155" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/156" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/157" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/158" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/159" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/160" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/161" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/162" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/163" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/164" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/165" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/166" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/167" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/168" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/169" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/170" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/171" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/172" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/173" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/174" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/175" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/176" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/177" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/178" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/179" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/180" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/181" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/182" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/183" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/184" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/185" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/186" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/187" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/188" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/189" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/190" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/191" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/192" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/193" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/194" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/195" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/196" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/197" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/198" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/199" />
<bp:OntologyReference rdf:resource="http://www.ebi.ac.uk/ontology-lookup/ontology/200" />
</bp:OntologyReference>
    
```



Aim: Convenient Access to Pathway Information



Facilitate creation and communication of pathway data
 Aggregate pathway data in the public domain
 Provide easy access for pathway analysis

Long term: Converge to integrated cell map



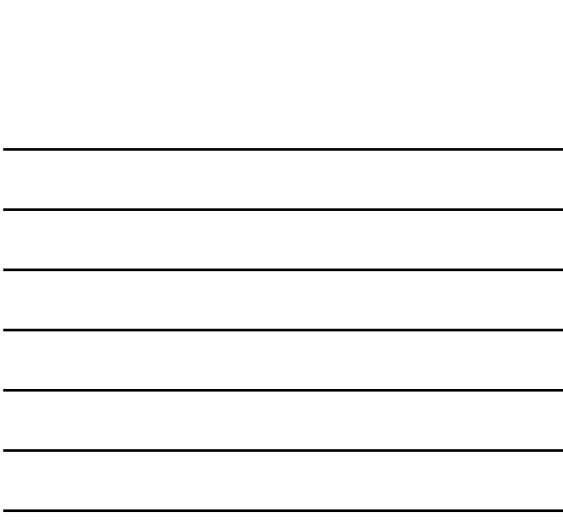
The screenshot shows the Pathway Commons homepage. The URL <http://pathwaycommons.org> is displayed vertically on the left. The page features a search bar at the top right and a navigation menu (Home, Filter, FAQ, About, Credits). The main content area includes a 'Pathway Commons Quick Stats' box with the following information:

- Number of Pathways: 921
- Number of Interactions: 5,924
- Number of Physical Entities: 15,515
- Number of Organisms: 10

Below the stats, there are sections for 'Biologists' (browse and search pathways) and 'Software developers' (build software on top of Pathway Commons). A 'Search Pathway Commons' section provides instructions on how to use the search function. At the bottom, a list of 'Pathway Commons currently contains the following data sources' is shown, including Cancer Cell Signaling, HumanCyc, NCI Nature Pathway Interaction Database, and Reactome.



The screenshot shows the search results for 'p53' on the Pathway Commons website. The search bar at the top contains 'p53'. The results page indicates that 22 relevant records were found. The 'Summary' section provides a detailed overview of the p53 pathway, including its role in DNA damage response and cell cycle regulation. The 'Data Sources' section lists the databases from which the pathways were retrieved. The 'Data Summary' section provides a list of pathways related to p53, such as 'Pathway: p53-Dependent G1/S DNA Damage Checkpoint' and 'Pathway: p53-Independent G1/S DNA Damage Checkpoint'. The URL <http://pathwaycommons.org> is displayed at the bottom of the page.



Pathway Commons Status

Pathway Commons Quick Stats:

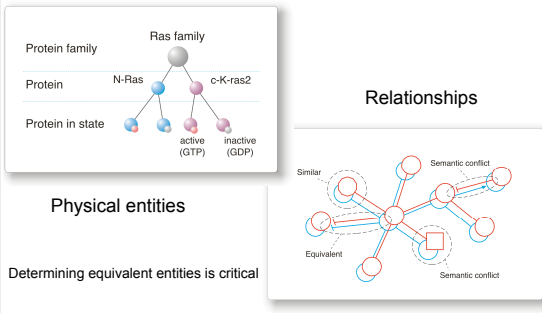
Number of Pathways: 1,449
 Number of Interactions: 421,395
 Number of Physical Entities: 88,509
 Number of Organisms: 441

- Signaling
- Metabolism
- Molecular Interactions
- Future
 - Genetic Interactions
 - Gene Regulation



Towards an Integrated Cell Map

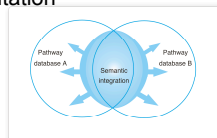
- Semantic pathway integration is difficult



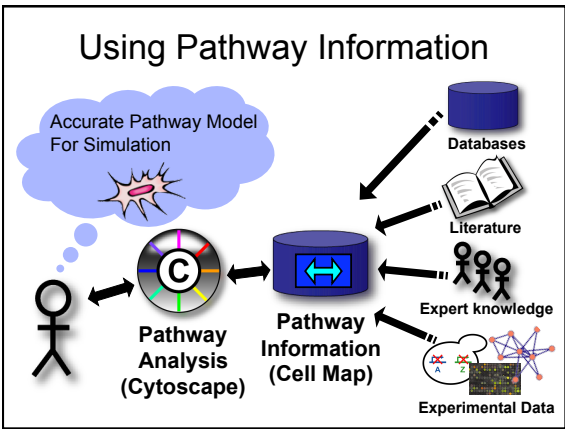
Practical Semantic Integration

- 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



Network Visualization and Analysis



Cytoscape

UNIVERSITY OF TORONTO

NCIBI
National Centre for Integrative Biomedical Informatics

Agilent Technologies

UCSF

Unilever

<http://cytoscape.org>

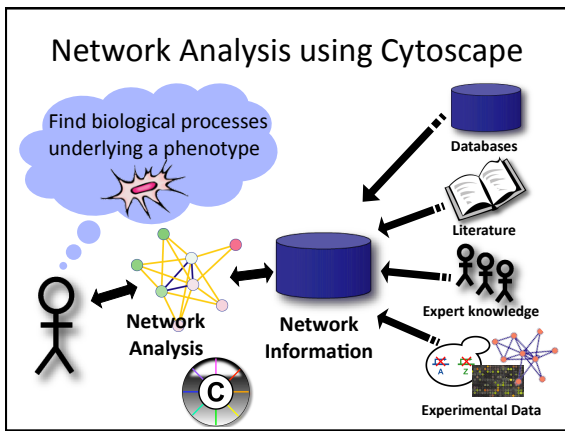
Network visualization and analysis

- Pathway comparison
- Literature mining
- Gene Ontology analysis
- Active modules
- Complex detection
- Network motif search

UCSD, ISB, Agilent, MSKCC, Pasteur, UCSF, Unilever, U Toronto, U Michigan

The screenshot displays the Cytoscape Desktop application. The main window shows a network graph with nodes and edges. A 'Node Attribute Browser' window is open, showing a table of node attributes:

Node	gplKno	gplKno	gplKno	gplKno	gplKno	gplKno
YCL08C	-0.332	-0.00976	-0.282	7.13664e-4	-0.379	3.26233e-2
YCL08FC	0.149	0.0012873	-0.085	0.11481	0.303	0.00277e-2
YNL45W	-0.304	0.148611	-0.098	0.09338	-1.237	0.109467
YML04W	-0.183	0.033372	-0.854	8.25344e-4	0.433	7.41121e-2



Manipulate Networks

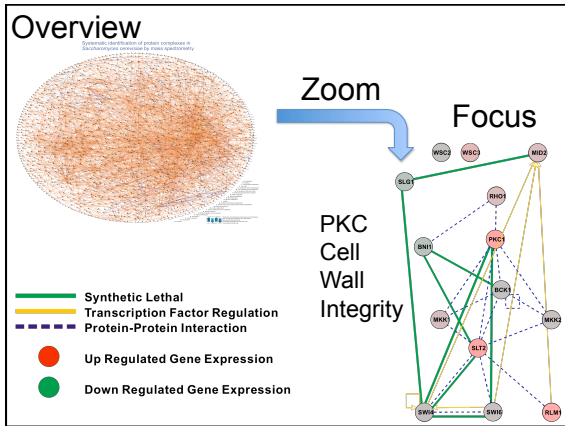
Filter/Query

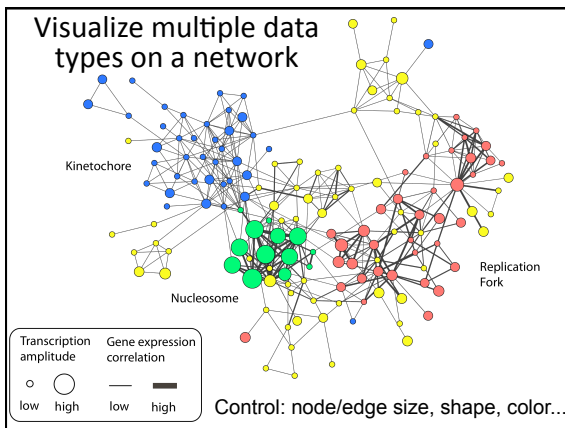
Automatic Layout

Interaction Database Search

This block contains four smaller screenshots of the Cytoscape interface:

- 'Manipulate Networks': Shows a network graph with various manipulation tools.
- 'Filter/Query': Shows the 'Control Panel' with a filter definition for 'node.Degree' ranging from 5 to 18.
- 'Automatic Layout': Shows a network graph with a circular layout.
- 'Interaction Database Search': Shows a search interface for the 'Cytoscape Network Protein Database'.





Active Community

<http://www.cytoscape.org>

- **Help**
 - 8 tutorials, >10 case studies
 - Mailing lists for discussion
 - Documentation, data sets
- **10,000s users, 2500 downloads/month**
- **>40 Plugins Extend Functionality**
 - Build your own, requires programming
 - e.g. Retina Workbench

Cline MS et al. Integration of biological networks and gene expression data using Cytoscape Nat Protoc. 2007;2(10):2366-82

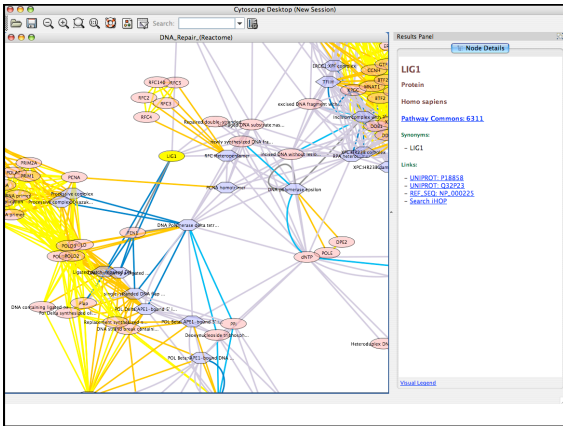
Analyzing Molecular Profiles

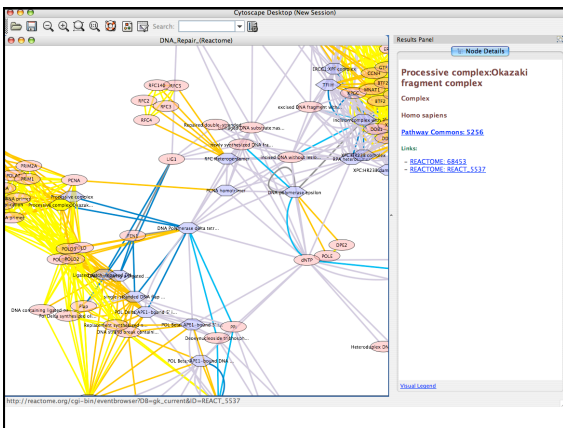
Analyzing gene expression data in a network context

- Input
 - Gene expression data
 - Network data
- Output
 - Visual diagram of expression data on network
 - Active network regions
- Outline
 - Where to find network data?
 - Interaction database (cPath)
 - Literature associations via text mining
 - Load expression data
 - Identify active pathways

Interaction Database Search

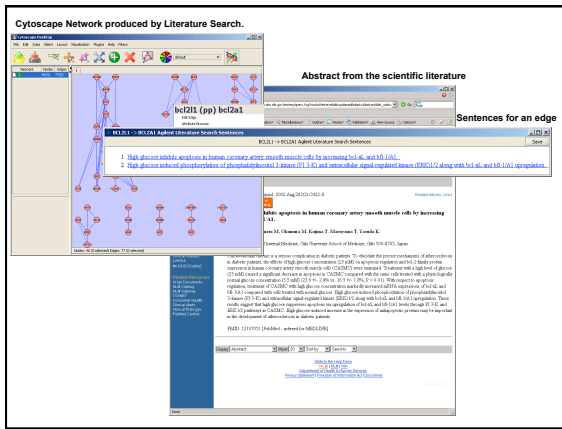
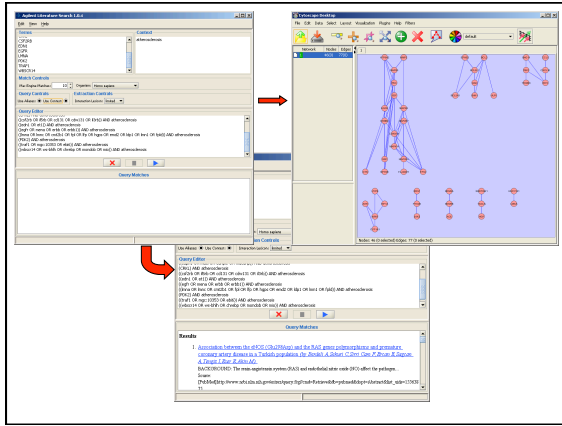
The screenshot displays the cPath software interface. On the left, a network visualization shows nodes connected by edges, with a 'Visual Legend' on the right. The main window shows the 'Import Network From Database' dialog box. The 'Data Source' is set to 'Pathway Commons Web Service Client'. The 'Search' field contains 'BRCA1'. The 'Species' dropdown is set to 'Human'. The 'Pathways' list includes 'BRCA1', 'ATM', 'RAD51', 'MDF1', 'MDF2', 'MDF3', 'MDF4', 'MDF5', 'MDF6', 'MDF7', 'MDF8', 'MDF9', 'MDF10', 'MDF11', 'MDF12', 'MDF13', 'MDF14', 'MDF15', 'MDF16', 'MDF17', 'MDF18', 'MDF19', 'MDF20', 'MDF21', 'MDF22', 'MDF23', 'MDF24', 'MDF25', 'MDF26', 'MDF27', 'MDF28', 'MDF29', 'MDF30', 'MDF31', 'MDF32', 'MDF33', 'MDF34', 'MDF35', 'MDF36', 'MDF37', 'MDF38', 'MDF39', 'MDF40', 'MDF41', 'MDF42', 'MDF43', 'MDF44', 'MDF45', 'MDF46', 'MDF47', 'MDF48', 'MDF49', 'MDF50', 'MDF51', 'MDF52', 'MDF53', 'MDF54', 'MDF55', 'MDF56', 'MDF57', 'MDF58', 'MDF59', 'MDF60', 'MDF61', 'MDF62', 'MDF63', 'MDF64', 'MDF65', 'MDF66', 'MDF67', 'MDF68', 'MDF69', 'MDF70', 'MDF71', 'MDF72', 'MDF73', 'MDF74', 'MDF75', 'MDF76', 'MDF77', 'MDF78', 'MDF79', 'MDF80', 'MDF81', 'MDF82', 'MDF83', 'MDF84', 'MDF85', 'MDF86', 'MDF87', 'MDF88', 'MDF89', 'MDF90', 'MDF91', 'MDF92', 'MDF93', 'MDF94', 'MDF95', 'MDF96', 'MDF97', 'MDF98', 'MDF99', 'MDF100'. The 'Details' section shows 'BRCA1' selected, with a list of associated pathways and a 'Load' button.





Text Mining

- Computationally extract gene relationships from text, usually PubMed abstracts
- Literature search tool, lots of network data
- BUT not perfect
 - Problems recognizing gene names
 - Natural language processing not perfect
- Agilent Literature Search Cytoscape plugin
- Others: E.g. iHOP
 - www.ihop-net.org/UniPub/iHOP/



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
- 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

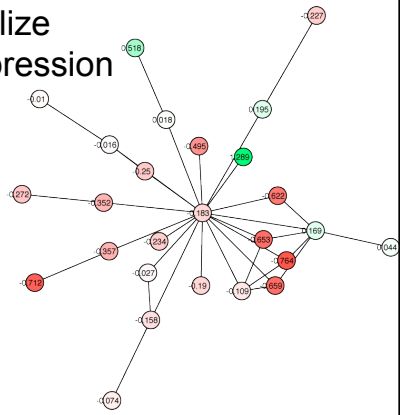
IDs to translate:

Output as spreadsheet:

	* entrezgene
YIL062C	854748
YLR370C	851085
YKL013C	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

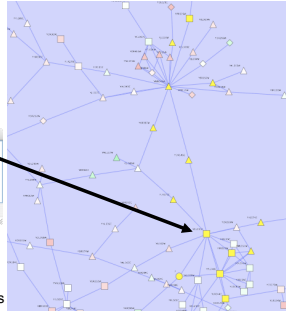
- Active modules
 - Input: network + p-values for gene expression values e.g. from GCRMA
 - Output: significantly differentially expressed subgraphs
- Method
 - Calculate z-score/node, Z_A score/subgraph, correct vs. random expression data sampling
 - Score over multiple experimental conditions
 - Simulated 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	
1	14	5.71	
2	26	3.584	
3	10	5.954	
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.

Bonus Slides

Gene and Protein Identifiers

- Identifiers (IDs) are names or numbers that help track database records
 - E.g. Social Insurance Number, Entrez Gene ID 41232
- Gene and protein information stored in many databases
 - → Genes have many IDs
- Records for: Gene, DNA, RNA, Protein
 - Important to use the correct record type
 - E.g. Entrez Gene records don't store sequence. They link to DNA regions, RNA transcripts and proteins.

Common Identifiers

Gene	Species-specific
Ensembl ENSG00000139618	HUGO HGNC BRCA2
Entrez Gene 675	MGI MGI:109337
Unigene Hs.34012	RGD 2219
	ZFIN ZDB-GENE-060510-3
RNA transcript	FlyBase CG9097
GenBank BC026160.1	WormBase WBGene00002299 or ZK1067.1
RefSeq NM_000059	SGD S000002187 or YDL029W
Ensembl ENST00000380152	Annotations
	InterPro IPR015252
Protein	OMIM 600185
Ensembl ENSP00000369497	Pfam PF09104
RefSeq NP_000050.2	Gene Ontology GO:0000724
UniProt BRCA2_HUMAN or	SNPs rs28897757
A1YBP1_HUMAN	Experimental Platform
IPI IPI00412408.1	Affymetrix 208368_3p_s_at
EMBL AF309413	Agilent A_23_P99452
PDB 1MIU	CodeLink GE60169
	Illumina GI_4502450-S
	Red = Recommended

ID Mapping Services

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: (NB: The strings in brackets are representative IDs in the corresponding namespace.)

Select "TO" namespace: (8541302)

File containing IDs to translate: [Browse...](#)

and/or

IDs to translate:

Output spreadsheet:

→

	entrezgene
Y1L062C	854748
YLR370C	851085
YKL013C	853856
YNR035C	855771
YBR234C	852536

- Synergizer
 - <http://lama.med.harvard.edu/cgi/synergizer/translate>
- Ensembl BioMart
 - <http://www.ensembl.org>
- PIR
 - <http://pir.georgetown.edu/pirwww/search/idmapping.shtml>

ID Mapping Challenges

- Gene name ambiguity
 - Not a good ID, but official gene symbol is ok e.g. HGNC/HUGO gene symbol
- Excel error-introduction
 - OCT4 is changed to October-4
- Problems reaching 100% coverage
 - E.g. due to version issues
 - Use multiple sources to increase coverage

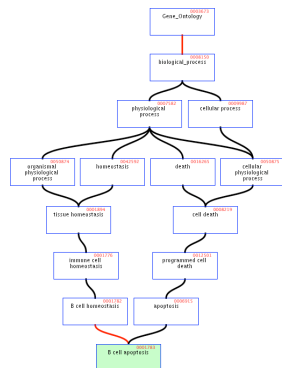
Zeeberg BR et al. Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics BMC Bioinformatics. 2004 Jun 23;5:80

Additional Plugins

- Bingo: over-representation analysis
- ClusterMaker: clusters networks, includes MCL
- NetworkAnalyzer: calculates statistics about a network
- (You may have to use an earlier version of Cytoscape to get some plugins to run)

The Gene Ontology (GO)

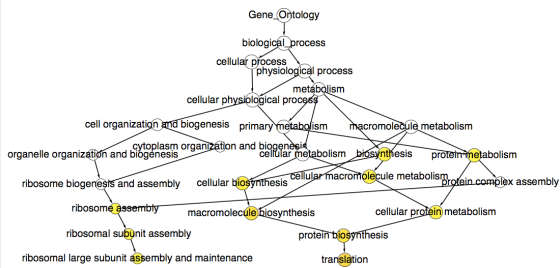
- Describes gene function
1. Agreed upon terms (controlled vocabulary)
 - Biological process
 - Cellular component
 - Molecular function
 2. Genome annotation



www.geneontology.org

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

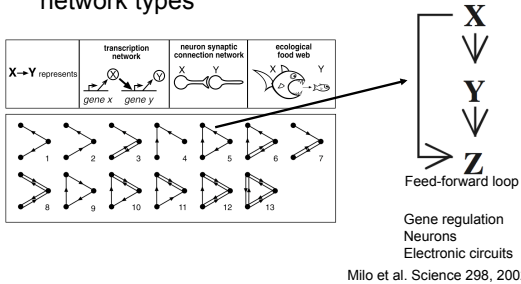
NetMatch

- 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>

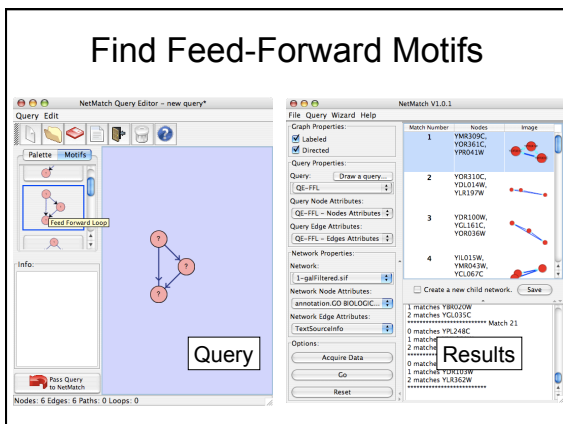
Ferro A, Giugno R, Pigola G, Pulvirenti A, Skripin D, Bader GD, Shasha D
 Bioinformatics 2007 Feb 3
 Extends state space representation based search from Cordella et al. IEEE
 Transactions on Pattern Analysis and Machine Intelligence, 2004, 26, 10, 1367–1372

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

Signaling pathway example

MAP Kinase Cascade
Nucleus - Growth Control Mitogenesis

NetMatch query

Shortest path between subgraph matches

NetMatch Results

Match Number	Nodes	Image
4	YL1157C, YMR043W, YLR229C	
5	YL1157C, YAL040C, YLR229C	
6	YLR100C, YER100W, YNL098C	

Find Expressed Motifs

Find specific subgraphs where certain nodes are significantly differentially expressed

NetMatch query

NetMatch Results

Protein	Differential Expression Significance
YLR075W	1.725E-4
YGR085C	2.639E-4
YPR102C	3.7183E-4

Systems Biology Graphical Notation

<http://sbgn.org>
