























Signaling	Pathway
1 2 3 4 5	
0	
gare 6A: The Cyclin - E2F cell cycle control system (version 3a - June 8, 19)	over.nci.nih.gov/kohnk/interaction_maps.htm













BUND

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























Interaction and Pathway Data Exchange Formats

- PSI-MI http://psidev.sourceforge.net
 Molecular interactions protein-protein interaction focus
 Peer reviewed, HUPO 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.celiml.org
 Math models of biological processes











BioPAX Pathway Language

- BioPAX = Biological Pathway Exchange
- Represent:
 - Metabolic pathways
 - Signaling pathways
 - Protein-protein, molecular interactions
 - Gene regulatory pathways
 - Genetic interactions
- Community effort: pathway databases distribute pathway information in standard format































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/



BioPAX Level 3 (in progress)

- · States and generics
 - E.g. phosphorylated P53, alcohols
- Gene regulation
 - E.g. Transcription regulation by transcription factors, translation regulation by miRNAs
- Genetic interactions
 E.g. synthetic lethality, epistasis
- Better controlled vocabulary integration
 More accessible to reasoners
- Switch to Protégé







• 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









The Ca	ncer Cell Map	Ethan Cerami Ben Gross
Memorial Sloa	en-Kattering Cancer Center	
Home Org	anisms FAQ Web Service Cytoscape About thway) from Homo sapiens	
Data	The Cancer Cell Map at Memorial Sloan-Kettering Cancer Center (http://cancer.cellmap.o	irg)
Source: Availability:	Freely available from The Cancer Cell Map (created by Memorial Sloan-Kettering Cancer	Center and the Institute of Bioinformatics) under the Creative
	Commons license (http://creativecommons.org/licenses/by/2.5/)	
View Express	EXPEGITE any intering of EXPT to the extractiluate domain of these EXTs likely to note times (EXTS extract), provines autophysiophysiothco, and ercentument of various signaling EXPT likels to the retruitment of various signaling proteins, including the safetyer proteins of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the wide potential of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extraction of the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra the extra	Larges (releva) in Ecuric pathway, it includes interactions or dimetration, activation of the infinite problem involvement of the second seco
Contains the	e Following Molecules (Showing 1 - 20 of 177) [display all]	
AKT1 AP2A1 A	APPL APPL2 ARAF1 ARF4 ATF1 BCAR1 CAMK2A CASP9 CAV1 CAV2 CBL CBLB	
CBL3 CDC42	CEACAM1 CEBPA CEBPB CREB	
Contains the	e Following Interactions / Pathways (Showing 1 - 20 of 475) [display all]	
Biochemical	I Reaction(s)	
View Details	EGFR → EGFR (Phosphorylated)	
View Details	ELK4 → ELK4 (Phosphorylated)	
View Details	GAP -> GAP (Phosphorylated)	
View Details	ERK5 (Phosphorylated) → ERK5	
View Details	MEKK3 -> MEKK3 (Phosphorylated)	
View Details	EPS15 → EPS15 (Phosphorylated)	
View Details	<u>JUN</u> → <u>JUN</u> (Phosphorylated)	
Miner Dataile	VAV2 VAV2 /Bleashandsted)	
cand	cer.cellmap.org	

-			





cPath Key Features

- · Identifier mapping system e.g. proteins
- Scalable pathway data aggregation
- · Simple web interface for browse and query
- Standard web service API for application communication
- 100% open source
 Java, Tomcat, MySQL, Lucene, Struts, YUI
- Local installation and customization

http://cbio.mskcc.org/cpath Cerami EG, Bader GD, Gross BE, Sander C. BMC Bioinformatics. 2006 Nov 13;7:497





























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







































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

















00	Filter Creation Dialog	Ne	etwork Fil	tering
-Filter types-	Filter Type Description			0
Numeric Filter	r Select nodes based on the			
String Filter	attributes of surrounding			
Topology Filte	er nodes			
Boolean Meta	-Filter			
Node Interact	tions			
e	OK	Use	Filters	
~~	Manage Filters		Node Topology Filter	
(Create new filter Remove selected f	ilter		
A	Available Filters		Filter Name	NodeTopology:>=1~1
N	lode Interaction:	0		
N	Node : canonicalName ~ *		Select nodes with at least	1 neighbors
E	dge : null<0.0	<u> </u>		
	Interaction:		within distance	1
E	dge Interaction:			
N	Node : Protein description ~ *phage*	-	that pass the filter	[No Filter]
N	<pre>iodeTopology:>=1~1</pre>	Ŧ	that pass the litter	
	Apply selected filter			



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







Text Mining

- Computationally extract gene relationships from text, usually PubMed abstracts
- · Literature search tool, lots of network data
- BUT not perfect
 - Problems recognizing gene namesNatural 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
 Ilama.med.harvard.edu/cgi/synergizer/translate
- More ID mapping services available
 http://baderlab.org/IdentifierMapping

(Gene E	xpr Inte	ess egra	ion/Ne ⁻ ation	two	rk	
THE SYN	ERGIZER						
The Synergizer data synonym relationshi naming scheme (a.l load sample inputs	abase is a growing repo ips. This tool facilitates k.a "namespace") to ar	ository of ge the conver- nother.	ene and pro sion of iden	tein identifier ifiers from one		ļ	
Select species:	Saccharomyces cerevisiae		•		*	entrezgene	
Select authority:	ensembl		-		YIL062C	854748	
Colore ISDOM			-	(MR). The steiner is	YKL013C	853856	
namespace:	ANY		•	[brackets] are	YNR035C	855771	
Select "TO" namespace:	entrezgene		[854192] •	representative IDs in the corresponding namespaces.)	YBR234C	852536	
File containing IDs to translate:		Browse					
and/or						-	
IDs to translate:	YIL062C YLR370C YKL013C YNR035C YBR234C			1. Load a Cytoscaj	as attribu pe	tes in	
Output as spreadsheet: Submit	Π			2. Assign expression values to nodes using this attribute set			





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













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















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





		,	ACODE Results Summary	
Rank	Score	Size	Names	Complex
1	7.25	16,116	YGR232W, YDL007W, YKL145W, YFR052W, YFR004W, YLR421C, YOR261C, YDL147W, YDR427W, YHR200W, YER021W, YOR117W, YDL097C, YOR259C, YPR108W, YDR394W	
2	6.387	31,198	VPL039W, VPL034W, V0R272W, VNL130C, VVL039W, VFL035C, VLA09C, VL129W, V1035C, VLA09C, VL129W, V0R061W, VR050W, VCR57C, VDR449C, V0R039W, VJL09C, VPL012W, VR0103W, VL09C, VPL012W, VRL014C, VLL08W, VRL172W, VNL002C, VRL02CW, VRL172W, VNL002C, VRL02CW, VRL172W, VNL002C, VRL032W, V0R145C, VPR016C	**
3	5.417	12,65	YGL011C, YOL038W, YPR103W, YMR314W, YBL041W, YOR362C, YER012W, YJL001W, YML092C, YGR253C, YER094C, YGR135W	
4	5	15,75	YPL043W, YMR290C, YER006W, YKR081C, YDR496C, YDL031W, YNL061W, YNL132W, YLR222C, YLR197W, YMR049C, YHR052W, YLL059C, YKL099C, YDL014W	
5	5	12,60	YPR187W, YPR010C, YPR110C, YNL248C, YOR341W, YNR003C, YKL144C, YOR207C, YPR190C, YNL113W, YOR116C, YBR154C	
	0	Create a new	child network. Save Done	
			Bader & Hogue, BMC Bi	oinformatics 2003 4(1)



⊖ ○ ⊖ MCODE Results Summary				
Rank	Score	Size	Names	Complex
1 Proteasome 2	7.25 26S	16,116	YGR232W, YDL007W, YKL145W, YFR052W, YFR004W, YLR421C, YOR261C, YDL147W, YDR427W, YHR200W, YER021W, YOR117W, YDL097C, YOR259C, YPR108W, YDR394W	
Ribosome 2	6.387	31,198	YEL039W, YEL004W, YOR272W, YNL110C, YKL009W, YFL002C, YOL077C, YPL126W, YIL035C, YLK409C, YLR129W, YOR061W, YKR060W, YCR057C, YDR449C, YOR039W, YJL109C, YPL012W, YKL014C, YLL005W, YKL172W, YNL002C, YCL111W, YOL041C, YCR002C, YCL111W, YOL041C, YCR002C, YCL111W, YOL041C, YCR002C, YCL111W,	**
Proteasome 2	205		YGL011C, YOL038W, YPR103W.	
T TOLEASOTTIE 2	.00		YMR314W, YBL041W, YOR362C,	
3	5.417	12,65	YER012W, YJL001W, YML092C, YGR253C, YER094C, YGR135W	
RNA Splicing	5	15,75	YPL043W, YMR290C, YER006W, YKR081C, YDR496C, YDL031W, YNL061W, YNL132W, YLR222C, YLR197W, YMR049C, YNL052W,	<u>R</u>
RNA Pol core	-	12.60	YJL069C, YRL099C, YDL014W YPR187W, YPR010C, YPR110C, YNL248C, YOR341W, YNR003C, YNL144C, YOR341W, YNR003C,	



Other Pathway Analyses

- Find known pathways in new species using Pathblast (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







