

BioPAX - Biological Pathway Data Exchange Format Tutorial

Gary Bader
CCBR, University of Toronto
BioPAX Workgroup
www.biopax.org
NETTAB June.12.2007.Pisa

BioPAX Supporting Groups

Current Participants

- Memorial Sloan-Kettering Cancer Center: E.Demir, M. Cary, C. Sander
- University of Toronto: G. Bader
- SRI Bioinformatics Research Group: P. Karp, S. Paley, J. Pick
- Bilkent University: U. Dogrusoz
- Université Libre de Bruxelles: C. Lemmerling
- CBRC Japan: K. Fukuda
- Dana Farber Cancer Institute: J. Zucker
- Millennium: J. Rees, A. Ruttenberg
- Cold Spring Harbor/EBI: G. Wu, M. Gillespie, P. D'Eustachio, I. Vastrik, L. Stein
- BioPathways Consortium: J. Luciano, E. Neumann, A. Regev, V. Schachter
- Argonne National Laboratory: N. Maltsev, E. Marland, M. Syed
- Harvard: F. Gibbons
- AstraZeneca: E. Pichler
- BIOBASE: E. Wingender, F. Schacherer
- NCI: M. Aladjem, C. Schaefer
- Università di Milano Bicocca, Pasteur, Rennes: A. Splendiani
- Vassar College: K. Dahlquist
- Columbia: A. Rzhetsky

Collaborating Organizations

- Proteomics Standards Initiative (PSI)
- Systems Biology Markup Language (SBML)
- CellML
- Chemical Markup Language (CML)

Databases

- BioCyc, WIT, KEGG, BIND, PharmGKB, aMAZE, INOH, Transpath, Reactome, PATIKA, eMIM, NCI PID, CellMap

Wouldn't be possible without

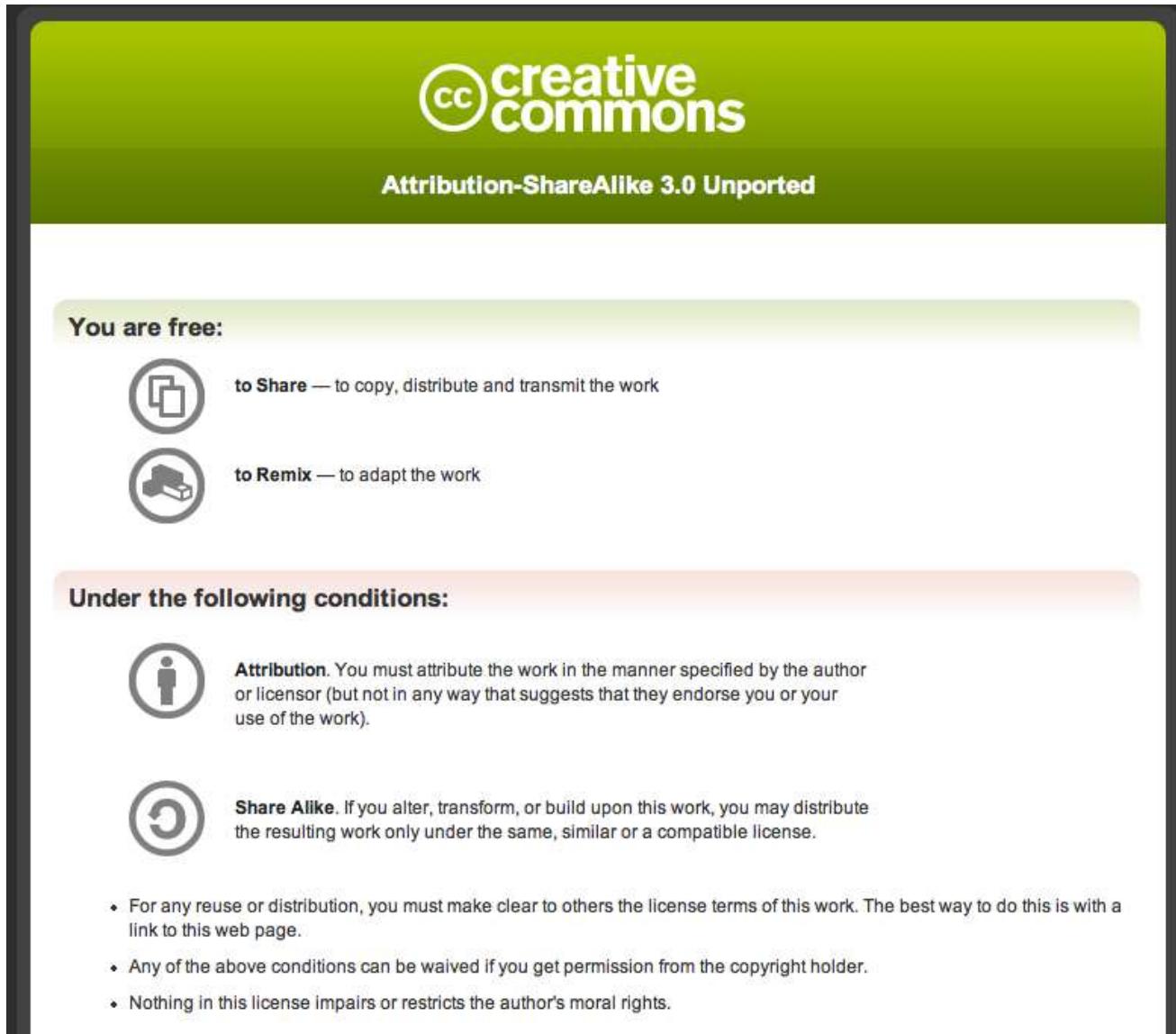
Gene Ontology

Protégé, U.Manchester, Stanford

Grants/Support

- Department of Energy (Workshop)
-  caBIG

<http://creativecommons.org/licenses/by-sa/3.0/>



The image shows the Creative Commons Attribution-ShareAlike 3.0 Unported license page. At the top, the Creative Commons logo is displayed next to the text "Attribution-ShareAlike 3.0 Unported". Below this, a green bar contains the text "You are free:" followed by two icons: one for sharing and one for remixing. Under the heading "Under the following conditions:", there are two main conditions: "Attribution" (represented by a person icon) and "Share Alike" (represented by a circular arrow icon). The "Attribution" condition requires attribution to the original author. The "Share Alike" condition requires that any derivative works must be distributed under the same or a compatible license. A bulleted list at the bottom provides additional details about the license terms.

You are free:

-  to Share — to copy, distribute and transmit the work
-  to Remix — to adapt the work

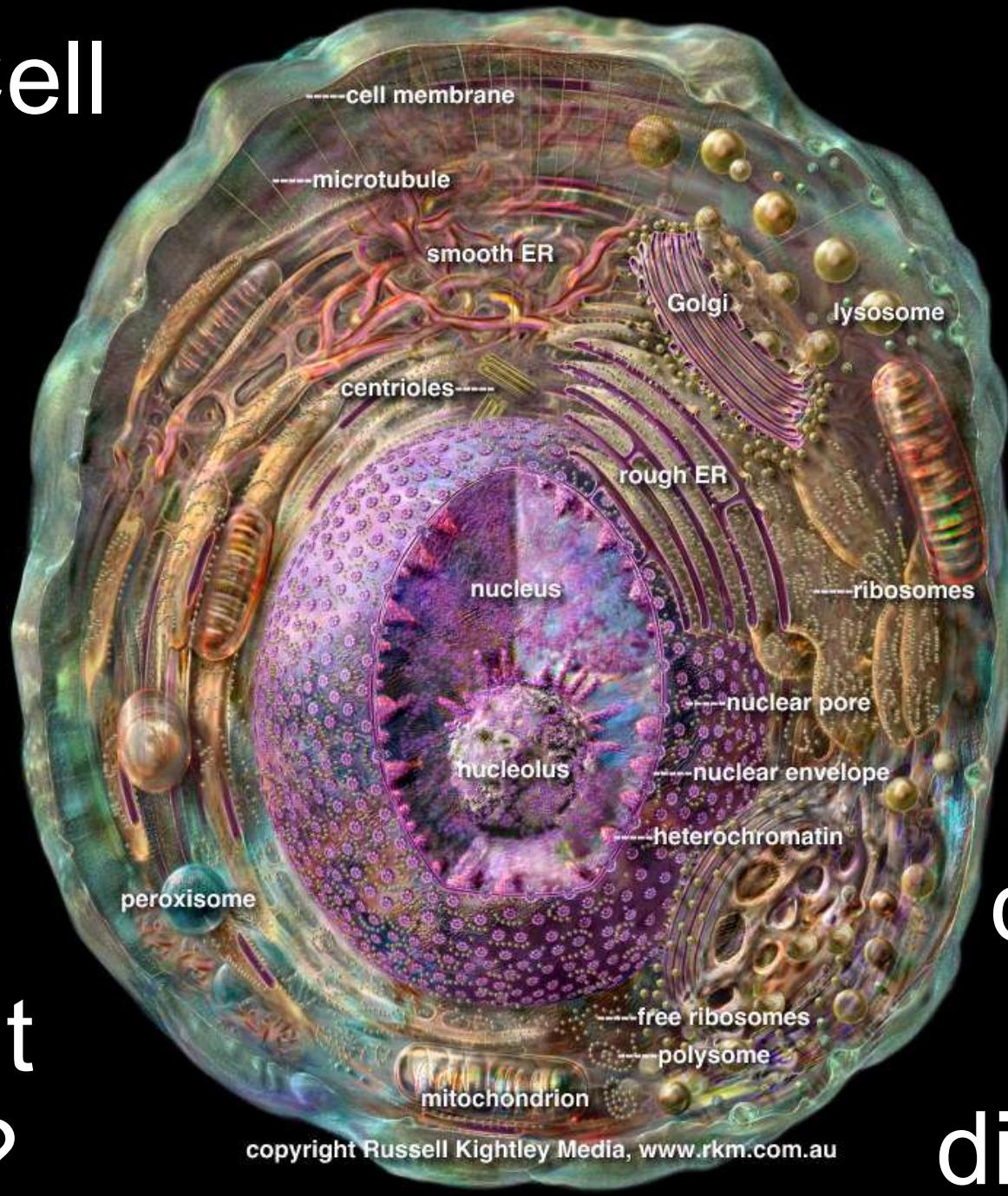
Under the following conditions:

-  **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, similar or a compatible license.
 - For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
 - Any of the above conditions can be waived if you get permission from the copyright holder.
 - Nothing in this license impairs or restricts the author's moral rights.

Will be made available from biopax.org wiki

The Cell

How
does it
work?



How
does it
fail in
disease?

Pathways

- Pathways are biological processes
- But, not really pathways → networks
- Metabolic, signaling, regulatory and genetic
- Define gene function at many different levels
- Biologists have found useful to group together for organizational, historic, biophysical or other reasons

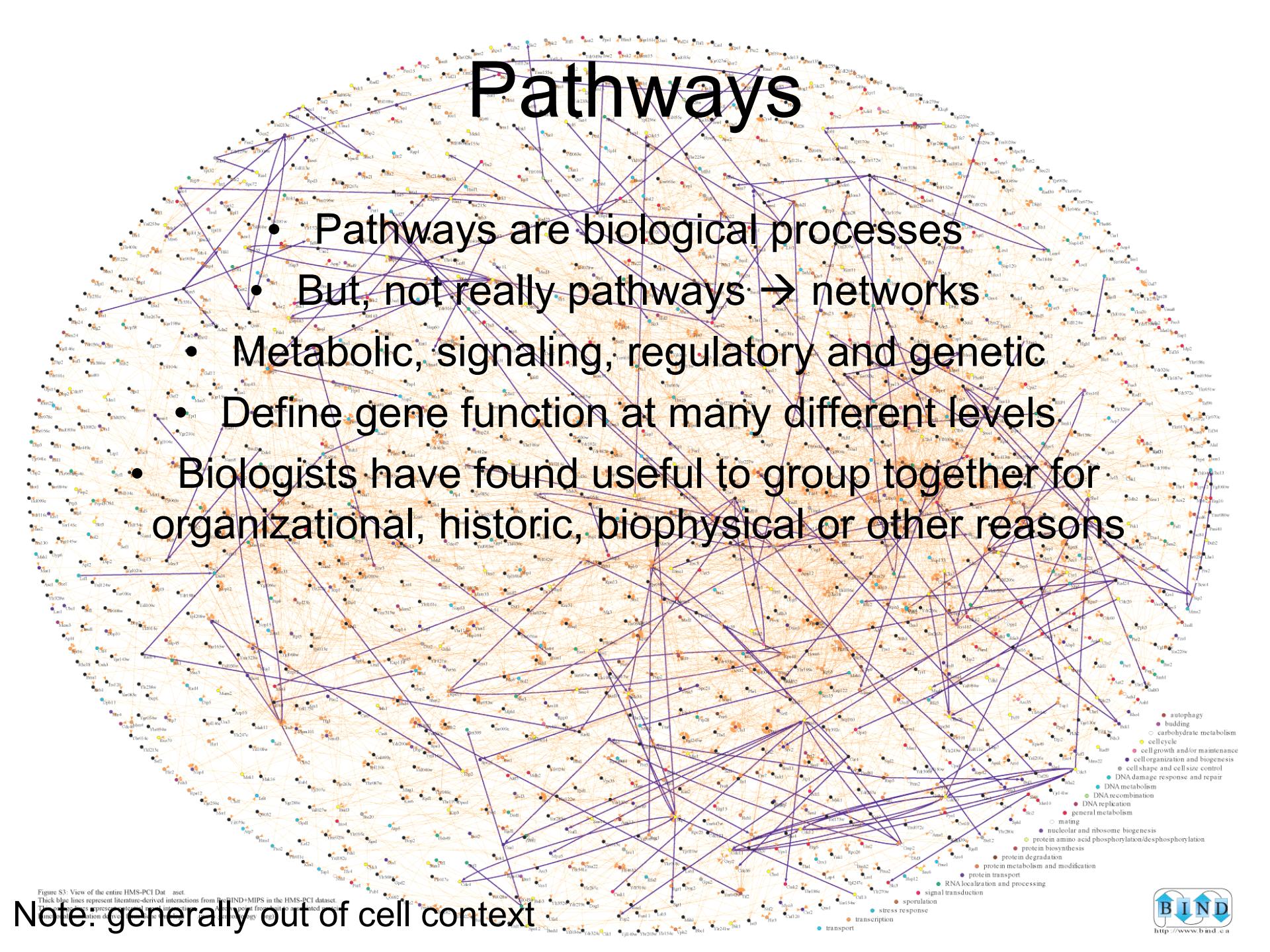
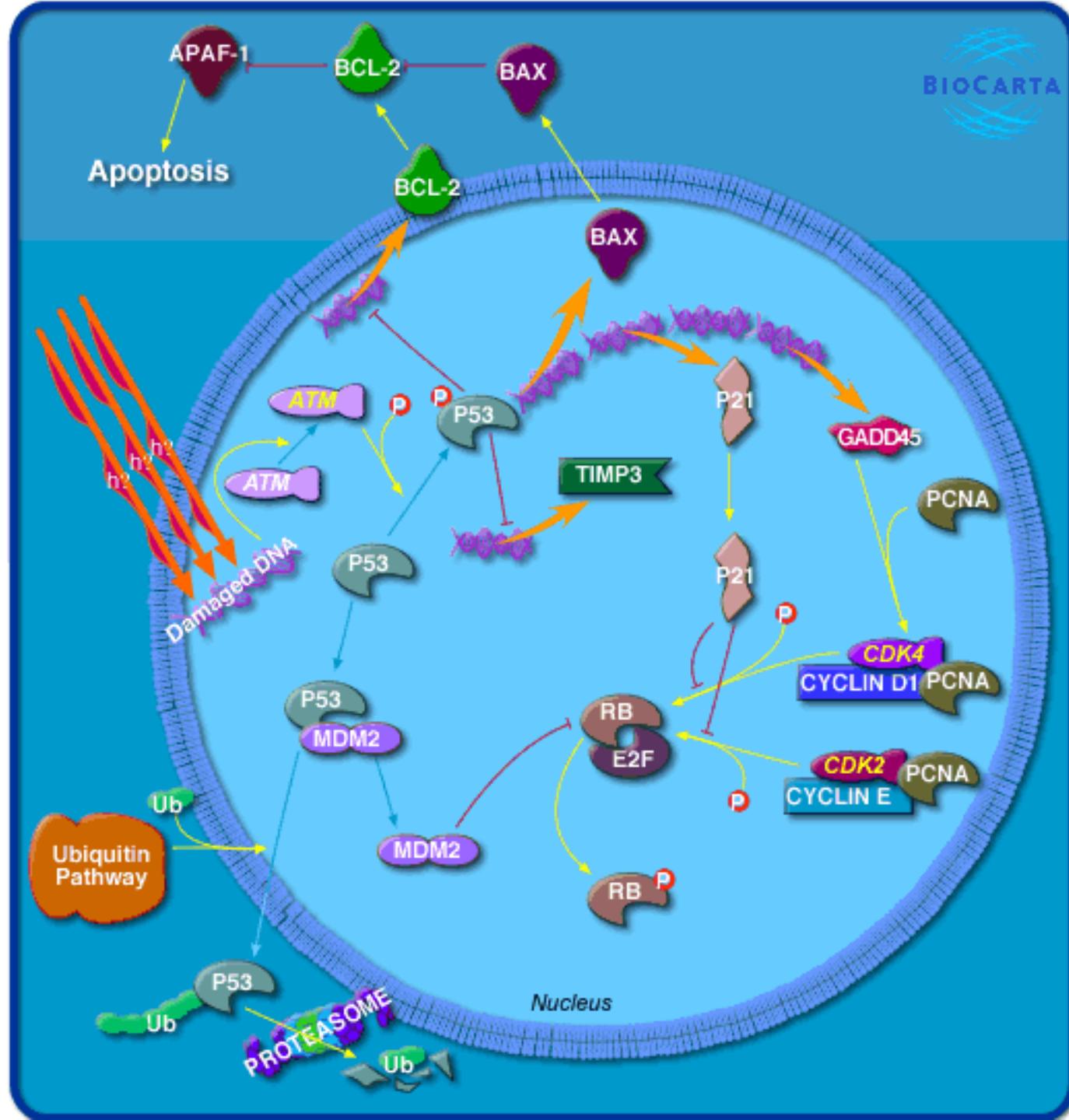
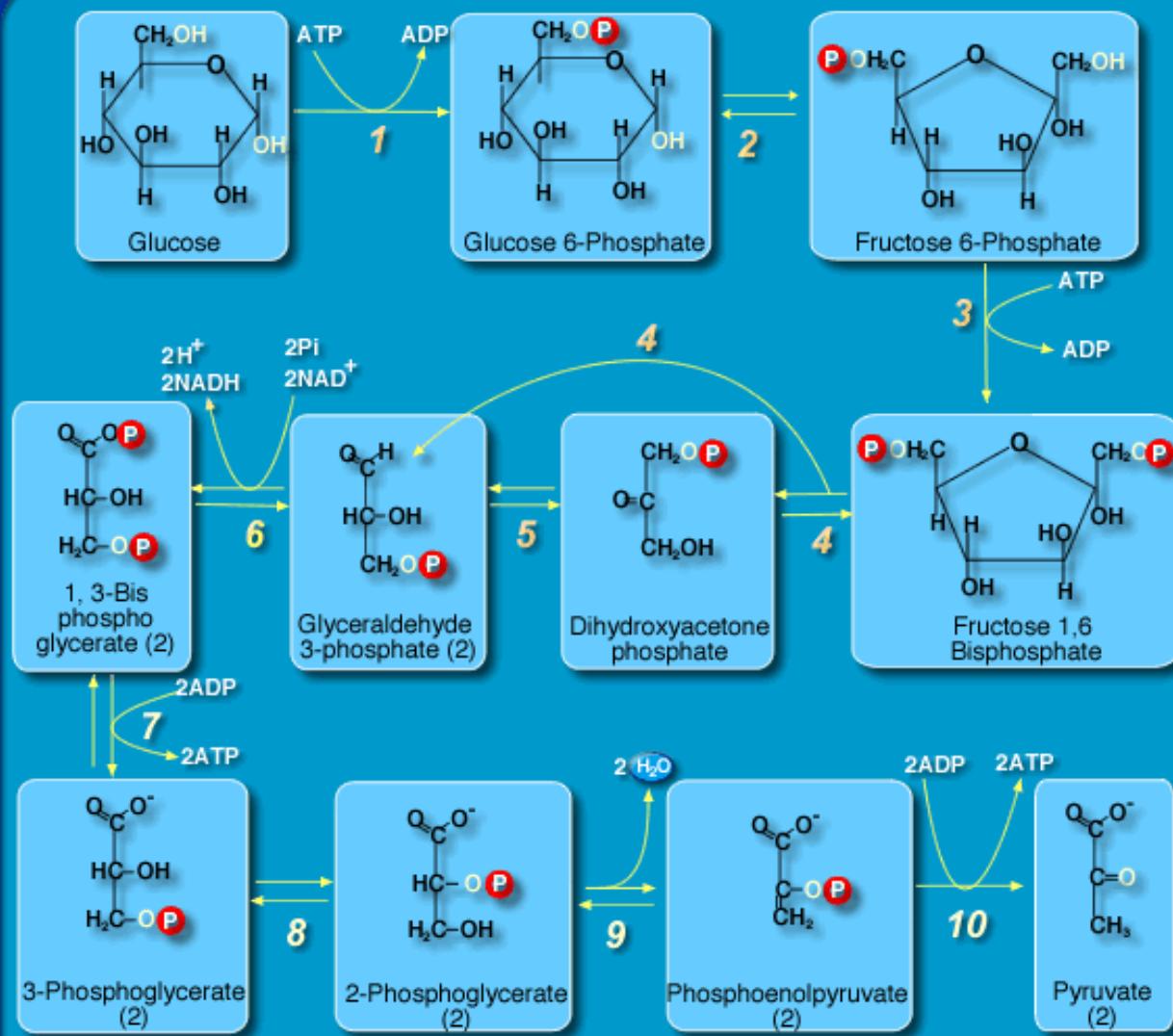


Figure S3: View of the entire HMS-PCI Data set.

Thick blue lines represent literature-derived interactions from BioGRID+MIPS in the HMS-PCI dataset. The nodes are represented by small colored circles. A node point from left to right indicates the literature-derived interaction type.

Note: generally out of cell context





ENZYMES

- 1 Hexokinase
- 2 Glucose Phosphate Isomerase
- 3 Phosphofructokinase
- 4 Fructose diphosphatase aldolase

● Preparatory phase

- 5 Triose phosphate Isomerase
- 6 Glyceraldehyde Phosphate Dehydrogenase

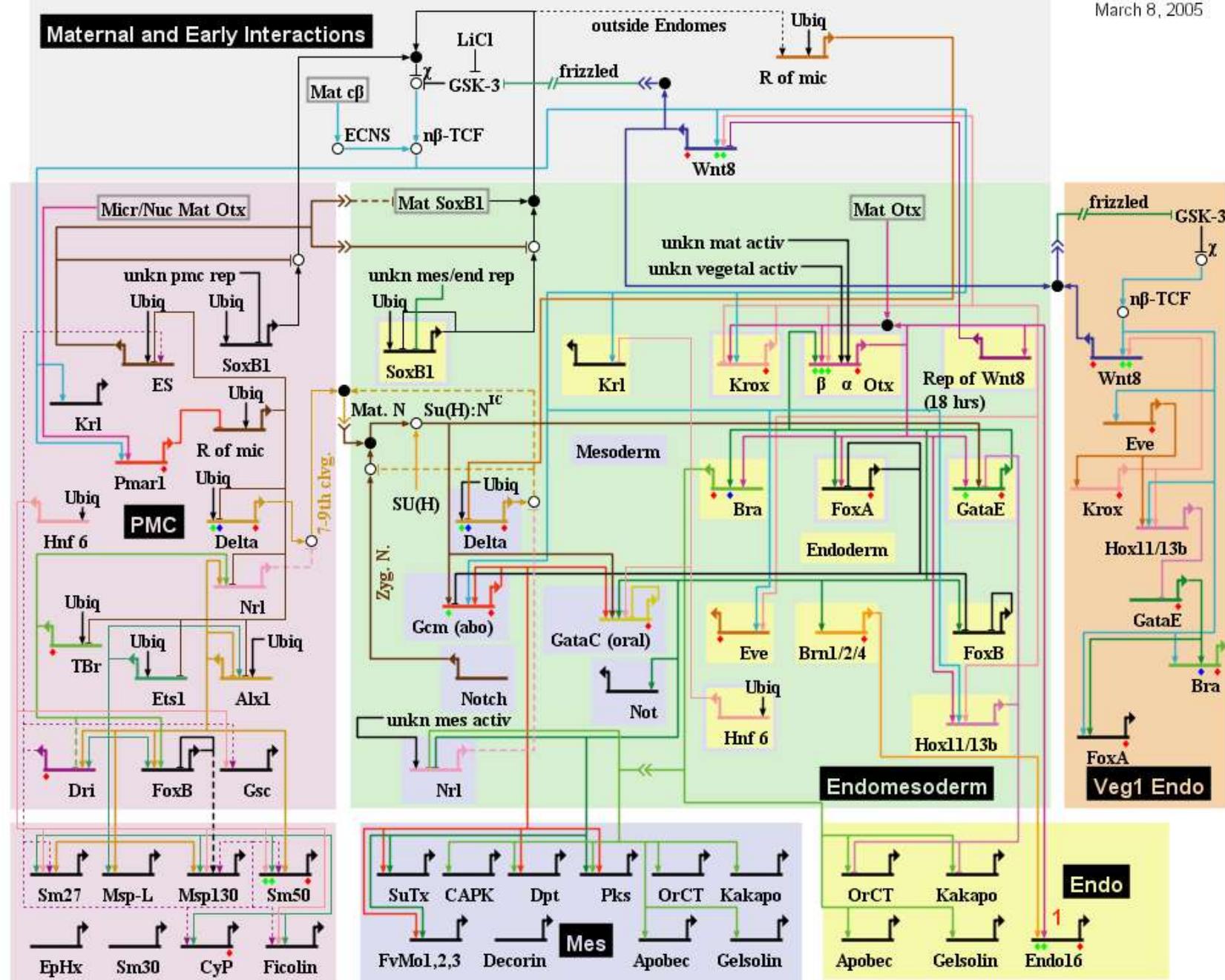
● Payoff phase

- 7 Phosphoglycerate Kinase
- 8 Phosphoglyceromutase
- 9 Enolase
- 10 Pyruvate Kinase

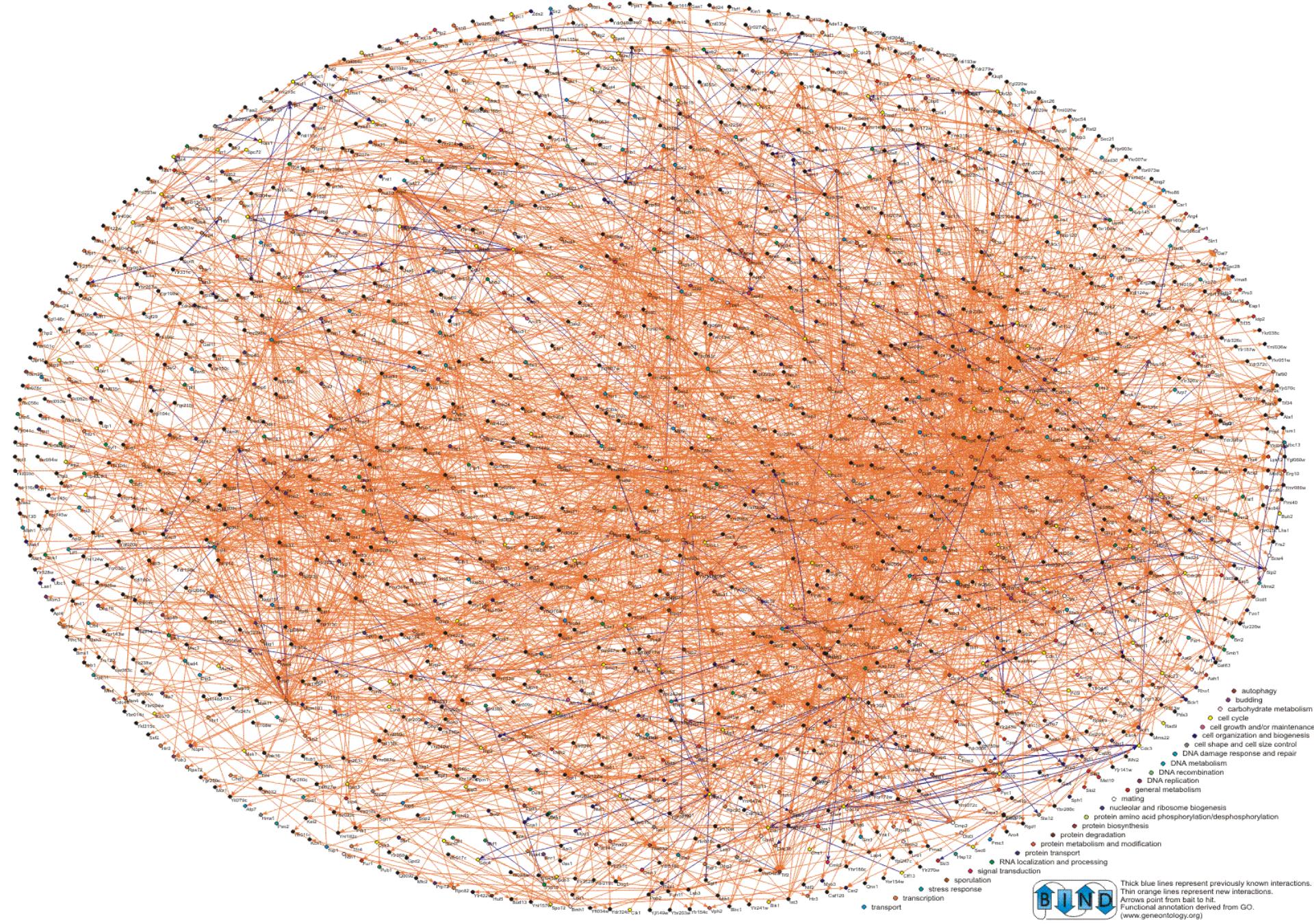


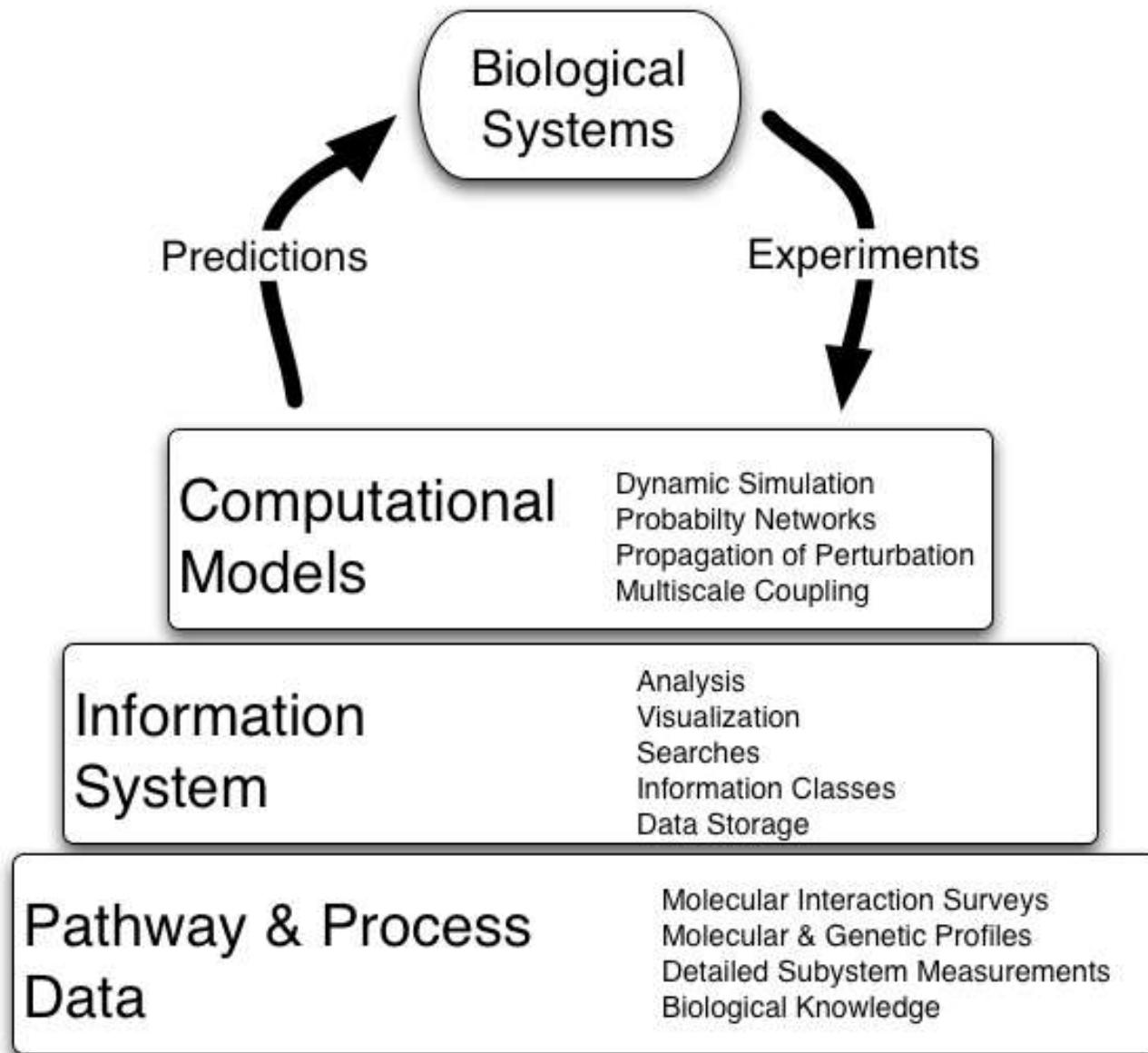
Endomesoderm Specification to 30 Hours

March 8, 2005



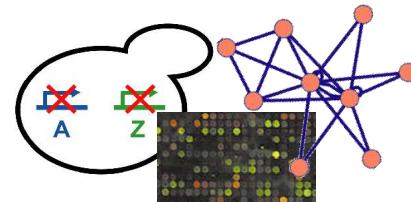
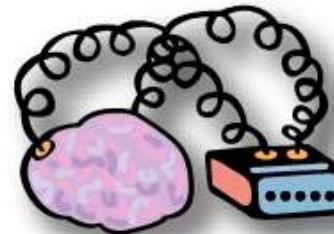
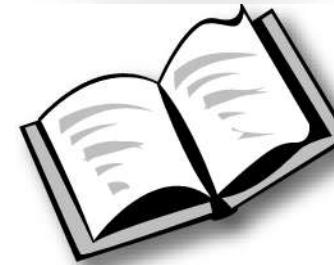
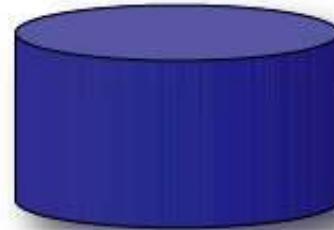
Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry





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



Pathway Databases

Metabolic Pathways aMAZE - Protein Function and Biochemical Pathways Project BioCyc - BioCyc Knowledge Library BioModels - BioModels Database BioSilico - BioSilico BRENDa - Comprehensive Enzyme Information System CellML Repository - CellML Model Repository EcoCyc - Encyclopedia of E. coli Genes and Metabolism EMP - Enzymes and Metabolic Pathways Database ENZYME - Enzyme nomenclature database GeneNet - Genetic Networks GenMAPP - Gene MicroArray Pathway Profiler GOLOdb - Genomics of Lipid-associated Disorders Indigo - Gene Neighborhoods and Colon Usage IntEnz - Integrated relational Enzyme database KEGG - Kyoto Encyclopedia of Genes and Genomes LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways Malaria - Malaria Parasite Metabolic Pathways MetaCore - MetaCore pathway database Metacyc - Metabolic Pathway Database MIPS CYGD - MIPS Comprehensive Yeast Genome Database Monod - Monod's Notebook and Datastore MPB - Metabolic Pathways of Biochemistry MRAD - Metabolic Reaction Analysis Database NetBioChem - Medical Biochemistry Resource Nicholson Minimaps - IUBMB-Nicholson Minimaps PathArt - Pathway Articulator PathDB - Pathways Database PATAKA - Pathway Analysis Tool for Integration and Knowledge Acquisition PDS - Pathways Database System PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base ProcesDB - ProcesDB Reactome - Reactome KnowledgeBase SBML Model Repository - SBML Model Repository Soybase - USDA-ARS Soybean Genetics and Genome Database TECR - Thermodynamics of Enzyme-Catalyzed Reactions UM-BD - Biocatalysis/Biodegradation Database WIT - What Is There?	BRIT - Biomolecular Relations in Information Transmission and Expression CTDB - Calmodulin Target Database DDB - Database of Domain Interactions and Bindings DIP - Database of Interacting Proteins Doodle - Database of oligomerization domains from lambda experiments DopAnet - DopAnet DRC - Database of Ribosomal Crosslinks DSM - Dynamic Signaling Maps FIMM - Functional Molecular Immunology FlyNets - FlyNets FusionDB - Prokaryote-Gene Fusion Events GPCR-PD - G protein-coupled receptors protein database GRID - General Repository for Interaction Datasets GroEL PPI - Proteins that interact with GroEL and factors that affect their release HIV-1 at NCBI - HIV-1, Human Protein Interaction Database at NCBI HIV-MD - HIV Molecular Immunology Database Hiv-DFI - Helicobacter pylori Database of Protein Interactomes HPI - Human Protein Interaction Database HPRD - Human Protein Reference Database HSV1 PPI - Protein-Protein Interactions Table for Human herpesvirus 1 HumanSD - Human Proteome Survey Database ICBS - Inter-Chain Beta-Sheets IntAct - IntAct INTERACT - INTERACT Protein-protein interaction database InterDom - Database of interacting Domains Interfaces - DATASET OF PROTEIN-PROTEIN INTERFACES Interolog - Interolog/Regulog Database JenPep - JenPep Peptide Binding database KDBI - Kinetic Data of Bio-molecular Interactions database KinaseDB - Kinase Pathway Database MHCPep - Database of MHC binding peptides MINT - Molecular Interactions Database MIPS CYGD - MIPS Comprehensive Yeast Genome Database MIRD - MHC Peptide Interaction Database MycoP@PDB - Human Fungal Pathogens Proteome Database NetPro - Molecular Connections NetPro ootTFD - Object Oriented Transcription Factors Database OPHID - The Online Predicted Human Interaction Database PathCatDB - Pathways Database PDZBase - PDZ domain protein-protein interaction database PhosphoELM - Post-translational phosphorylation database PhosphoBase - Database of phosphorylation sites PhosphoSite - Cell Signaling Technology's PhosphoSite Database PIBASE - PIBASE PIMDb - Drosophila Protein Interaction Map Database PIMer - Protein Interaction Map - Hybrigenics PINdb - Proteins Interacting in the Nucleus database POINT - Prediction of interactome PombePD - Schizosaccharomyces pombe Proteome Database	PPD - Protein-Protein Interaction Database PPV (FANTOM) - RIKEN FANTOM Protein Protein Interaction Viewer POS - Protein Quaternary Structure database Predictome - Predictome ProChart - ProChart database of signal transduction pathway information ProLinks - ProLinks ProMesh - ProMesh Protein-Protein Interaction Database ProNet - Protein-protein Interaction Database	CNDB - Cell Signaling Networks Database DOOGS - Database of Quantitative Cellular Signaling DSM - Dynamic Signaling Maps eMIM - Electronic Molecular Interaction Map GeneNet - Genetic Networks GenMAP - Gene Microarray Pathway Profiler GOLDdb - Genomics of Lipid-associated Disorders GON - Genomic Object Net INOH - Integrating Network Objects with Hierarchies	STCDB - Signal Transduction Classification Database STKE - Signal Transduction Knowledge Environment TRANSPATH - Signal Transduction Browser TRMP - Therapeutically Relevant Multiple Pathways Database TRRD - Transcription Regulatory Regions Database XPD - Phosphorylation Site Database	CTDB - Comparative Toxicogenomics Database HetPDB_Nav1 - HetPDB Nav1 KIBank - KIBank Ligand Depot - Ligand Depot MDR - Metalloprotein Database NRR - Nuclear Receptor Resource ORDB - Olfactory Receptor Database PDB-Ligand - PDB-Ligand PDSP - Psychoactive Drug Screening Program Kit Database PLD - Protein Ligand Database Relibase - Protein-ligand database query tool ResNet - ResNet TTD - Therapeutic Target Database
220 Pathway Databases!					
Protein-Protein Interactions 3D-3D interacting domains ABCdb - Archaea and Bacteria ABC transporter database AFCS - Alliance for Cellular Signaling Molecule Pages Database AIFuse - Functional Associations of Proteins in Complete Genomes ASEdb - Alanine Scanning Energies Database ASPD - Artificial Selected Proteins/Peptides Database BID - Binding Interface Database BIND - Biomolecular Interaction Network Database BindingDB - The Binding Database	iSLbase - Protein Structural Interactome Database PubGene - PubGene RedNet - RedNet	MetaCore - MetaCore pathway database Monod - Monod's Notebook and Datastore PANTHER - PANTHER	PathArt - Pathway Articulator PathDB - Pathways Database Pathways Knowledge Base - Ingenuity Pathways Knowledge Base Acquisition PDS - Pathways Database System PhosphoSite - Cell Signaling Technology's PhosphoSite Database PID - CMAP Pathway Interaction Database ProcessDB - ProcesDB Reactome - Reactome KnowledgeBase RedNet - RedNet	JASPAR - JASPAR Transcription Factor Binding Profile Database MAPPER - MAPPER ootTFD - Object Oriented Transcription Factors Database PRODRIC - PROkaryotic database of gene regulation RegulonDB - Database on Transcriptional Regulation and Genome Organization SCPD - The Promoter Database of Saccharomyces cerevisiae TRACTOR - TRACTOR_D8 TRANSFAC - Transcription Factor Database TRED - Transcriptional Regulatory Element Database TRRD - Transcription Regulatory Regions Database	Protein-Compound Interactions CIEBE - Computed Ligand Binding Energy Other SELEX_DB - Randomized DNA/RNA sequence database AARSDB - Aminoacyl-tRNA Synthetase Database MedGene - MedGene AACNT - Amino Acid-Nucleotide Interaction Database ProNIT - Thermodynamic Database for Protein-Nucleic Acid Interactions TCDB - Transport Classification Database TransportDB - TransportDB DPIDB - DNA-Protein Interaction Database IHOP - Information Hyperlinked Over Proteins PRID - Protein-RNA Interaction Database Precise - Predicted and Consensus Interaction Sites in Enzymes
AKCS - Alliance for Cellular Signaling Molecule Pages Database aMAZE - Protein Function and Biochemical Pathways Project BioCyc - BioCyc Knowledge Library BioModels - BioModels Database BioSilico - BioSilico BRENDa - Comprehensive Enzyme Information System CellML Repository - CellML Model Repository EcoCyc - Encyclopedia of E. coli Genes and Metabolism EMP - Enzymes and Metabolic Pathways Database ENZYME - Enzyme nomenclature database GeneNet - Genetic Networks GenMAPP - Gene MicroArray Pathway Profiler GOLOdb - Genomics of Lipid-associated Disorders Indigo - Gene Neighborhoods and Colon Usage IntEnz - Integrated relational Enzyme database KEGG - Kyoto Encyclopedia of Genes and Genomes LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways Malaria - Malaria Parasite Metabolic Pathways MetaCore - MetaCore pathway database Metacyc - Metabolic Pathway Database MIPS CYGD - MIPS Comprehensive Yeast Genome Database Monod - Monod's Notebook and Datastore MPB - Metabolic Pathways of Biochemistry MRAD - Metabolic Reaction Analysis Database NetBioChem - Medical Biochemistry Resource Nicholson Minimaps - IUBMB-Nicholson Minimaps PathArt - Pathway Articulator PathDB - Pathways Database PATAKA - Pathway Analysis Tool for Integration and Knowledge Acquisition PDS - Pathways Database System PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base ProcesDB - ProcesDB Reactome - Reactome KnowledgeBase SBML Model Repository - SBML Model Repository Soybase - USDA-ARS Soybean Genetics and Genome Database TECR - Thermodynamics of Enzyme-Catalyzed Reactions UM-BD - Biocatalysis/Biodegradation Database WIT - What Is There?	AKCS - Alliance for Cellular Signaling Molecule Pages Database aMAZE - Protein Function and Biochemical Pathways Project BioCyc - BioCyc Knowledge Library BioModels - BioModels Database BioSilico - BioSilico BRENDa - Comprehensive Enzyme Information System CellML Repository - CellML Model Repository EcoCyc - Encyclopedia of E. coli Genes and Metabolism EMP - Enzymes and Metabolic Pathways Database ENZYME - Enzyme nomenclature database GeneNet - Genetic Networks GenMAPP - Gene MicroArray Pathway Profiler GOLOdb - Genomics of Lipid-associated Disorders Indigo - Gene Neighborhoods and Colon Usage IntEnz - Integrated relational Enzyme database KEGG - Kyoto Encyclopedia of Genes and Genomes LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways Malaria - Malaria Parasite Metabolic Pathways MetaCore - MetaCore pathway database Metacyc - Metabolic Pathway Database MIPS CYGD - MIPS Comprehensive Yeast Genome Database Monod - Monod's Notebook and Datastore MPB - Metabolic Pathways of Biochemistry MRAD - Metabolic Reaction Analysis Database NetBioChem - Medical Biochemistry Resource Nicholson Minimaps - IUBMB-Nicholson Minimaps PathArt - Pathway Articulator PathDB - Pathways Database PATAKA - Pathway Analysis Tool for Integration and Knowledge Acquisition PDS - Pathways Database System PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base ProcesDB - ProcesDB Reactome - Reactome KnowledgeBase SBML Model Repository - SBML Model Repository Soybase - USDA-ARS Soybean Genetics and Genome Database TECR - Thermodynamics of Enzyme-Catalyzed Reactions UM-BD - Biocatalysis/Biodegradation Database WIT - What Is There?	AKCS - Alliance for Cellular Signaling Molecule Pages Database aMAZE - Protein Function and Biochemical Pathways Project BioCyc - BioCyc Knowledge Library BioModels - BioModels Database BioSilico - BioSilico BRENDa - Comprehensive Enzyme Information System CellML Repository - CellML Model Repository EcoCyc - Encyclopedia of E. coli Genes and Metabolism EMP - Enzymes and Metabolic Pathways Database ENZYME - Enzyme nomenclature database GeneNet - Genetic Networks GenMAPP - Gene MicroArray Pathway Profiler GOLOdb - Genomics of Lipid-associated Disorders Indigo - Gene Neighborhoods and Colon Usage IntEnz - Integrated relational Enzyme database KEGG - Kyoto Encyclopedia of Genes and Genomes LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways Malaria - Malaria Parasite Metabolic Pathways MetaCore - MetaCore pathway database Metacyc - Metabolic Pathway Database MIPS CYGD - MIPS Comprehensive Yeast Genome Database Monod - Monod's Notebook and Datastore MPB - Metabolic Pathways of Biochemistry MRAD - Metabolic Reaction Analysis Database NetBioChem - Medical Biochemistry Resource Nicholson Minimaps - IUBMB-Nicholson Minimaps PathArt - Pathway Articulator PathDB - Pathways Database PATAKA - Pathway Analysis Tool for Integration and Knowledge Acquisition PDS - Pathways Database System PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base ProcesDB - ProcesDB Reactome - Reactome KnowledgeBase SBML Model Repository - SBML Model Repository Soybase - USDA-ARS Soybean Genetics and Genome Database TECR - Thermodynamics of Enzyme-Catalyzed Reactions UM-BD - Biocatalysis/Biodegradation Database WIT - What Is There?	AKCS - Alliance for Cellular Signaling Molecule Pages Database aMAZE - Protein Function and Biochemical Pathways Project BioCyc - BioCyc Knowledge Library BioModels - BioModels Database BioSilico - BioSilico BRENDa - Comprehensive Enzyme Information System CellML Repository - CellML Model Repository EcoCyc - Encyclopedia of E. coli Genes and Metabolism EMP - Enzymes and Metabolic Pathways Database ENZYME - Enzyme nomenclature database GeneNet - Genetic Networks GenMAPP - Gene MicroArray Pathway Profiler GOLOdb - Genomics of Lipid-associated Disorders Indigo - Gene Neighborhoods and Colon Usage IntEnz - Integrated relational Enzyme database KEGG - Kyoto Encyclopedia of Genes and Genomes LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways Malaria - Malaria Parasite Metabolic Pathways MetaCore - MetaCore pathway database Metacyc - Metabolic Pathway Database MIPS CYGD - MIPS Comprehensive Yeast Genome Database Monod - Monod's Notebook and Datastore MPB - Metabolic Pathways of Biochemistry MRAD - Metabolic Reaction Analysis Database NetBioChem - Medical Biochemistry Resource Nicholson Minimaps - IUBMB-Nicholson Minimaps PathArt - Pathway Articulator PathDB - Pathways Database PATAKA - Pathway Analysis Tool for Integration and Knowledge Acquisition PDS - Pathways Database System PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base ProcesDB - ProcesDB Reactome - Reactome KnowledgeBase SBML Model Repository - SBML Model Repository Soybase - USDA-ARS Soybean Genetics and Genome Database TECR - Thermodynamics of Enzyme-Catalyzed Reactions UM-BD - Biocatalysis/Biodegradation Database WIT - What Is There?	AKCS - Alliance for Cellular Signaling Molecule Pages Database aMAZE - Protein Function and Biochemical Pathways Project BioCyc - BioCyc Knowledge Library BioModels - BioModels Database BioSilico - BioSilico BRENDa - Comprehensive Enzyme Information System CellML Repository - CellML Model Repository EcoCyc - Encyclopedia of E. coli Genes and Metabolism EMP - Enzymes and Metabolic Pathways Database ENZYME - Enzyme nomenclature database GeneNet - Genetic Networks GenMAPP - Gene MicroArray Pathway Profiler GOLOdb - Genomics of Lipid-associated Disorders Indigo - Gene Neighborhoods and Colon Usage IntEnz - Integrated relational Enzyme database KEGG - Kyoto Encyclopedia of Genes and Genomes LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways Malaria - Malaria Parasite Metabolic Pathways MetaCore - MetaCore pathway database Metacyc - Metabolic Pathway Database MIPS CYGD - MIPS Comprehensive Yeast Genome Database Monod - Monod's Notebook and Datastore MPB - Metabolic Pathways of Biochemistry MRAD - Metabolic Reaction Analysis Database NetBioChem - Medical Biochemistry Resource Nicholson Minimaps - IUBMB-Nicholson Minimaps PathArt - Pathway Articulator PathDB - Pathways Database PATAKA - Pathway Analysis Tool for Integration and Knowledge Acquisition PDS - Pathways Database System PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base ProcesDB - ProcesDB Reactome - Reactome KnowledgeBase SBML Model Repository - SBML Model Repository Soybase - USDA-ARS Soybean Genetics and Genome Database TECR - Thermodynamics of Enzyme-Catalyzed Reactions UM-BD - Biocatalysis/Biodegradation Database WIT - What Is There?	AKCS - Alliance for Cellular Signaling Molecule Pages Database aMAZE - Protein Function and Biochemical Pathways Project BioCyc - BioCyc Knowledge Library BioModels - BioModels Database BioSilico - BioSilico BRENDa - Comprehensive Enzyme Information System CellML Repository - CellML Model Repository EcoCyc - Encyclopedia of E. coli Genes and Metabolism EMP - Enzymes and Metabolic Pathways Database ENZYME - Enzyme nomenclature database GeneNet - Genetic Networks GenMAPP - Gene MicroArray Pathway Profiler GOLOdb - Genomics of Lipid-associated Disorders Indigo - Gene Neighborhoods and Colon Usage IntEnz - Integrated relational Enzyme database KEGG - Kyoto Encyclopedia of Genes and Genomes LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways Malaria - Malaria Parasite Metabolic Pathways MetaCore - MetaCore pathway database Metacyc - Metabolic Pathway Database MIPS CYGD - MIPS Comprehensive Yeast Genome Database Monod - Monod's Notebook and Datastore MPB - Metabolic Pathways of Biochemistry MRAD - Metabolic Reaction Analysis Database NetBioChem - Medical Biochemistry Resource Nicholson Minimaps - IUBMB-Nicholson Minimaps PathArt - Pathway Articulator PathDB - Pathways Database PATAKA - Pathway Analysis Tool for Integration and Knowledge Acquisition PDS - Pathways Database System PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base ProcesDB - ProcesDB Reactome - Reactome KnowledgeBase SBML Model Repository - SBML Model Repository Soybase - USDA-ARS Soybean Genetics and Genome Database TECR - Thermodynamics of Enzyme-Catalyzed Reactions UM-BD - Biocatalysis/Biodegradation Database WIT - What Is There?

- Arguably the most accessible data source, but...
 - Varied formats, representation, coverage
 - Pathway data extremely difficult to combine and use
- Pathguide Pathway Resource List (<http://www.pathguide.org>)**

Navigation

- [Protein-Protein Interactions](#)
- [Metabolic Pathways](#)
- [Signaling Pathways](#)
- [Pathway Diagrams](#)
- [Transcription Factors / Gene Regulatory Networks](#)
- [Protein-Compound Interactions](#)
- [Genetic Interaction Networks](#)
- [Protein Sequence Focused](#)
- [Other](#)

Search

Organisms	<input type="button" value="All"/>
Availability	<input type="button" value="All"/>
Standards	<input type="button" value="All"/>
<input type="button" value="Reset"/>	<input type="button" value="Search"/>

Statistics[Analyze Pathguide](#)**Contact**

Comments, Questions,
Suggestions are
Always Welcome!

Complete Listing of All Pathguide Resources

Pathguide contains information about **222** biological pathway resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI 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](#).

News

Find it Easier
Many new search options are available

Get the Stats
Detailed Pathguide resource statistics now available

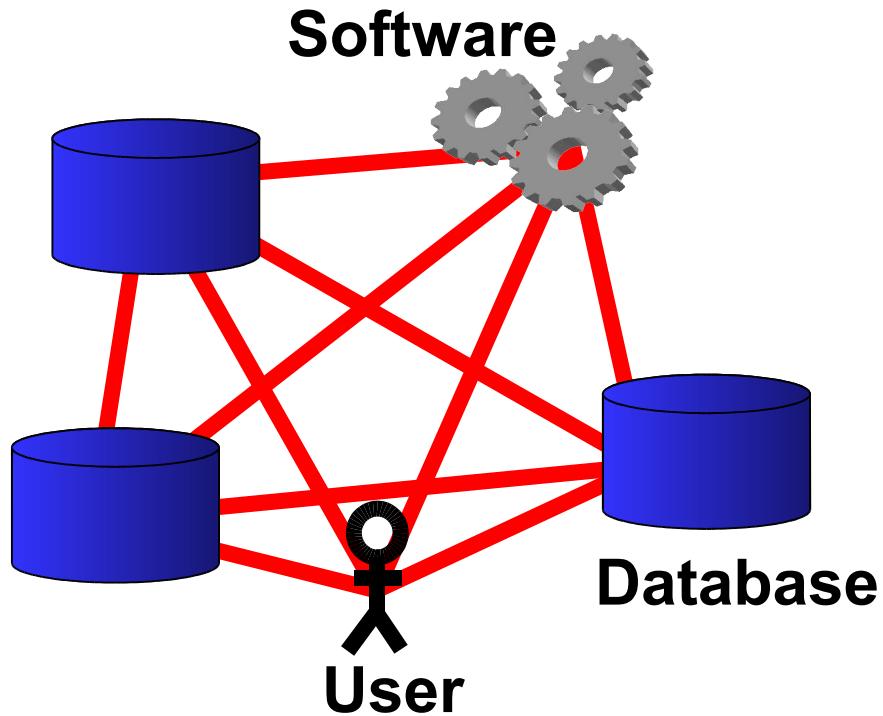
Pathguide Published
Please cite the [Pathguide](#)

Protein-Protein Interactions

Database Name (Order: alphabetically | [by web popularity](#))

		Full Record	Availability	Standards
3DID - 3D interacting domains		Details	 Free	
ABCdb - Archaea and Bacteria ABC transporter database		Details	 Free	
AfCS - Alliance for Cellular Signaling Molecule Pages Database		Details	 Free	
AllFuse - Functional Associations of Proteins in Complete Genomes		Details	 Free	
ASEdb - Alanine Scanning Energetics Database		Details	 Free	
ASPD - Artificial Selected Proteins/Peptides Database		Details	 ?	
BID - Binding Interface Database		Details	 Free	
BIND - Biomolecular Interaction Network Database		Details	 Free	 PSI-MI
BindingDB - The Binding Database		Details	 Free	
BioGRID - General Repository for Interaction Datasets		Details	  PSI-MI	 PSI-MI
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	 BioPAX
CSP - Cytokine Signaling Pathway Database		Details	 Free	
CTDB - Calmodulin Target Database		Details	 Free	
DDIB - Database of Domain Interactions and Bindings		Details	 Free	
DIP - Database of Interacting Proteins		Details	  PSI-MI	 PSI-MI
Doodle - Database of oligomerization domains from lambda experiments		Details	 Free	
DopaNet - DopaNet		Details	 Free	
DRC - Database of Ribosomal Crosslinks		Details	 Free	
DSM - Dynamic Signaling Maps		Details	 \$	
FIMM - Functional Molecular Immunology		Details	Free	
FusionDB - Prokaryote Gene Fusion Events		Details	Free	

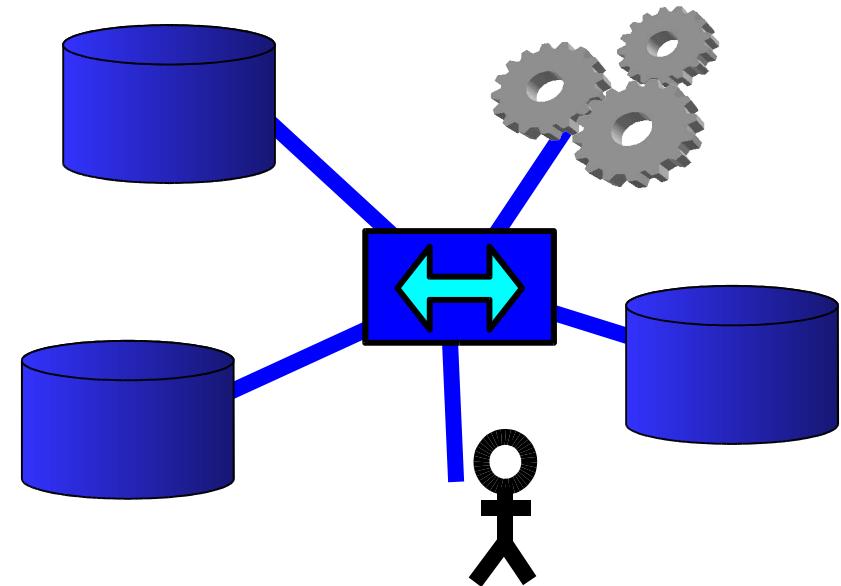
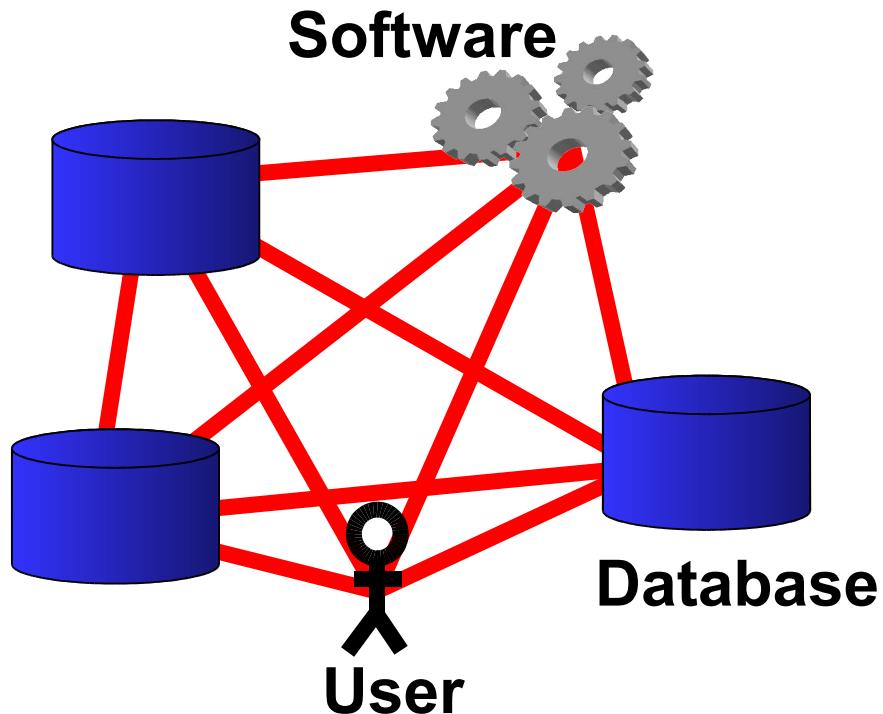
Gathering Pathway Information is Hard



>100 DBs and tools

Tower of Babel

Biological Pathway Exchange (BioPAX)



Before BioPAX

>100s DBs and tools

Tower of Babel

After BioPAX

Unifying language

Reduces work, promotes collaboration, increases accessibility

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

Ontologies: Components

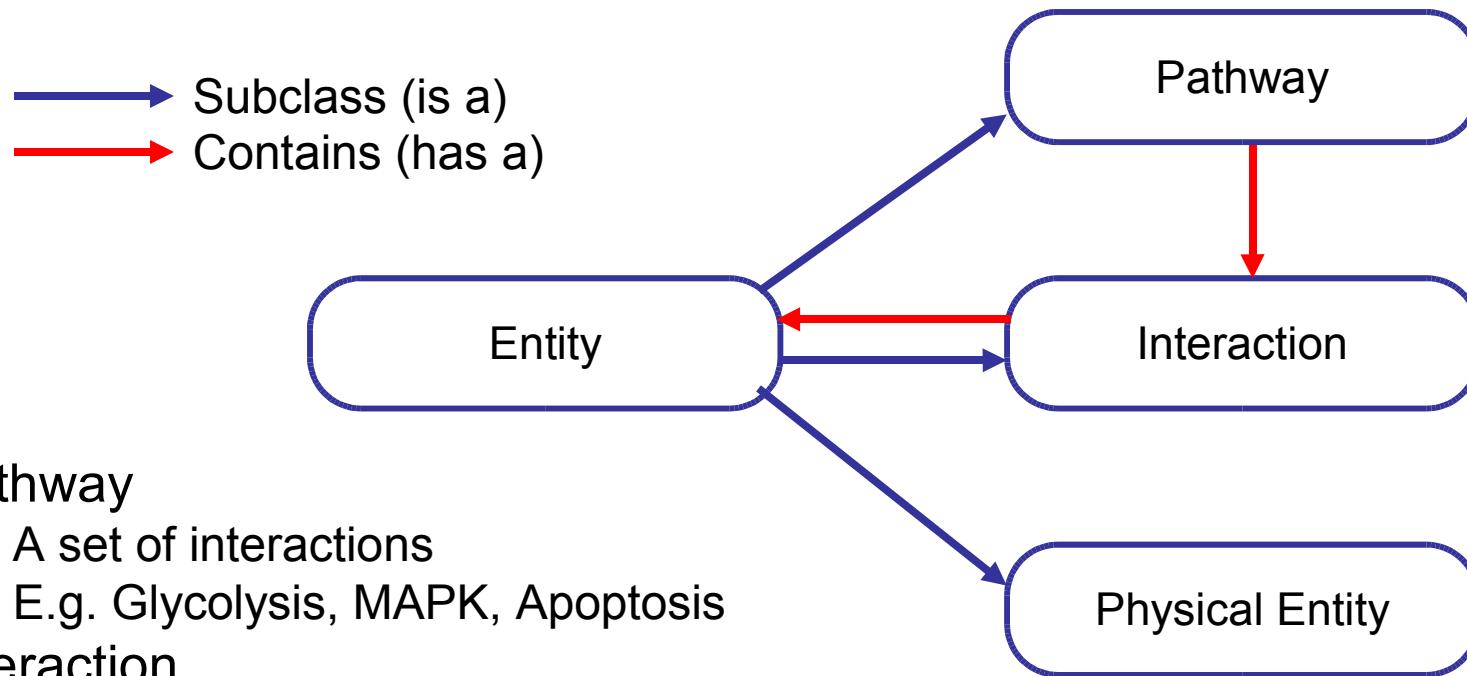
- Classes, relations & attributes, constraints, objects, values
- Classes (AKA “Concepts”, “Types”)
 - Arranged into a specialization hierarchy (AKA “Taxonomy”)
 - Parent-child relationships between classes
 - Class A is a parent of class B iff all instances of B are also instances of A
 - E.g. “Protein”, “RNA”, “Reaction”
- Relations & Properties (AKA “Slots”, “Attributes”, “Fields”)
 - Classes *have* properties, which may have values of specific types
 - Relationships: the value type is some other class in the ontology
 - E.g. “Substrate”, “Transporter”, “Participant”
 - Attributes: the value type is a simple data type
 - E.g. “Molecular Wt.”, “Sequence”, “ ΔG ”

From Peter Karp, “Ontologies: Definitions, Components, Subtypes”, SRI International, presentation available at <http://www.biopax.org>

Ontologies: Components (cont)

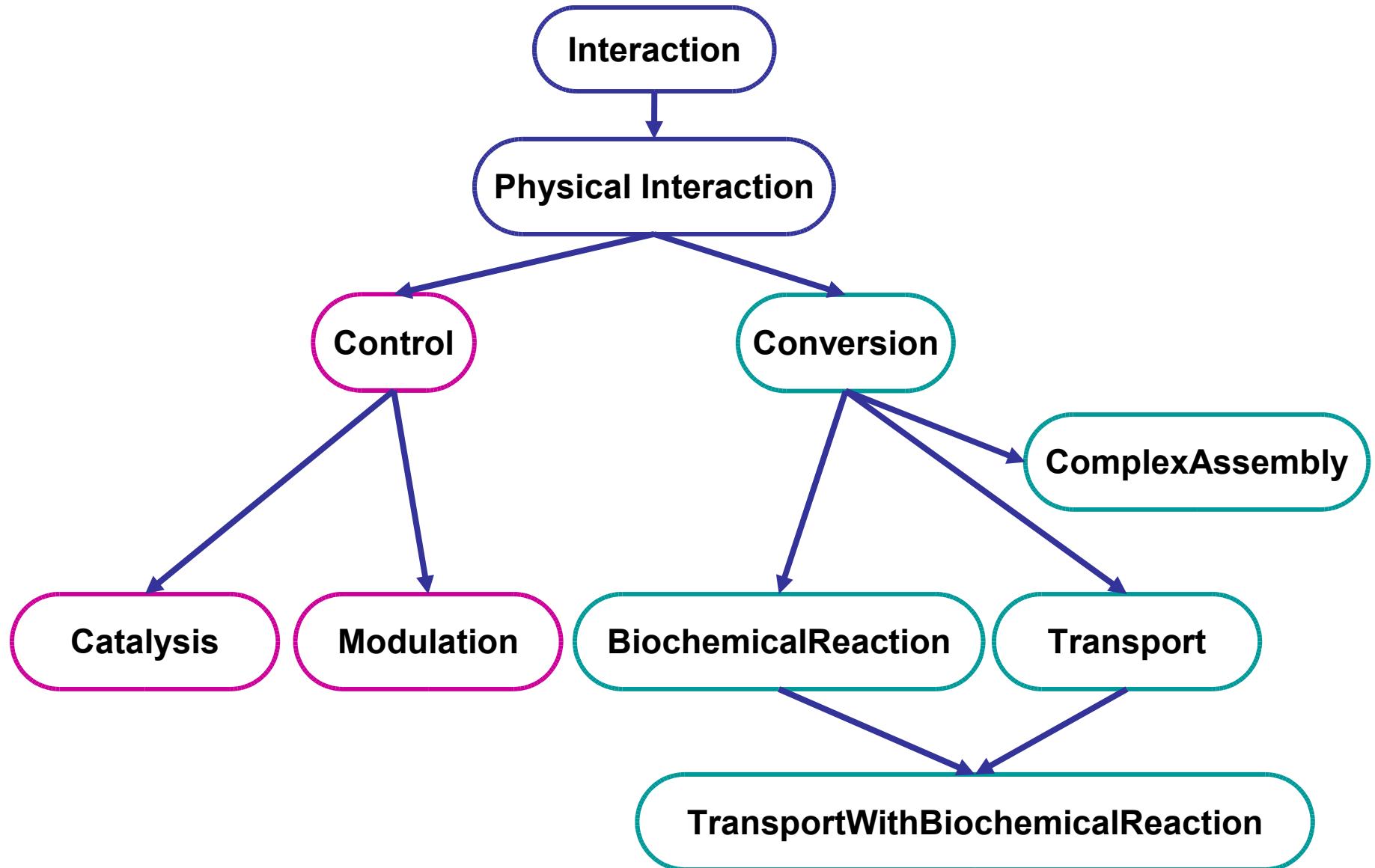
- Constraints
 - Define allowable values and connections within an ontology
 - E.g. “MOLECULAR_WT must be a positive real number”
- Objects and Values
 - Objects are instances of classes
 - Values occupy the slots of those instances
 - Strictly speaking, an ontology with instances is a knowledge base
 - Beyond the scope of BioPAX workgroup, our users will create the instances of classes in the BioPAX ontology

BioPAX Structure

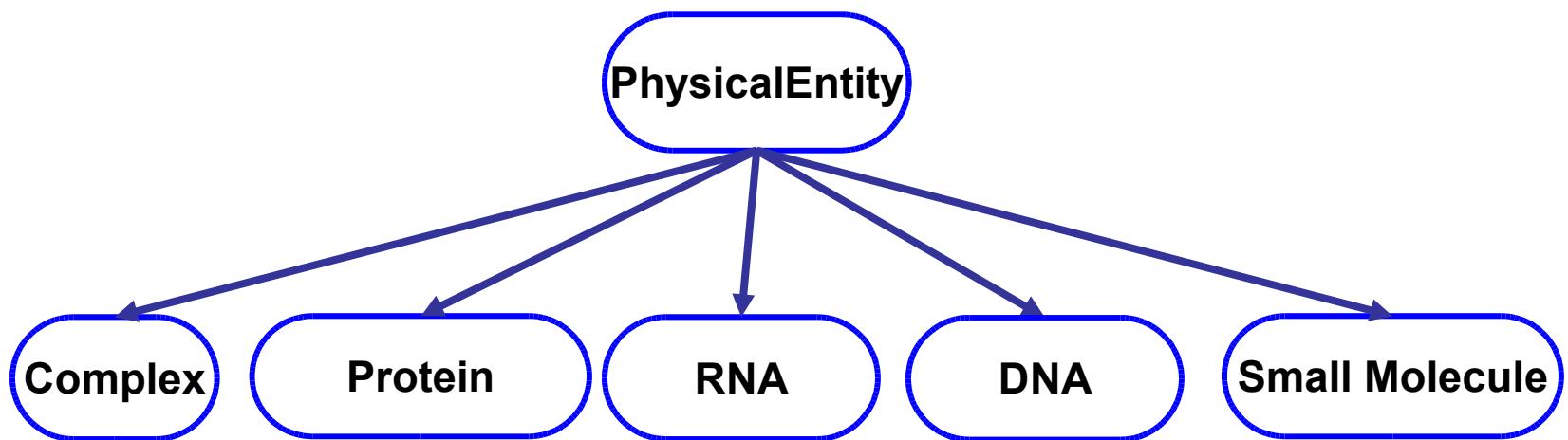


- **Pathway**
 - A set of interactions
 - E.g. Glycolysis, MAPK, Apoptosis
- **Interaction**
 - A basic relationship between a set of entities
 - E.g. Reaction, Molecular Association, Catalysis
- **Physical Entity**
 - A building block of simple interactions
 - E.g. Small molecule, Protein, DNA, RNA

BioPAX: Interactions



BioPAX: Physical Entities



BioPAX Ontology

conversion	
0	COMMENT
0	PARTICIPANTS
0	AVAILABILITY
0	DATA-SOURCE
>	SYNONYMS
0	SHORT-NAME
0	NAME
0	XREF
0	RIGHT
0	LEFT
0	SPONTANEOUS

control	
0	COMMENT
0	PARTICIPANTS
0	AVAILABILITY
0	DATA-SOURCE
0	SYNONYMS
0	SHORT-NAME
0	NAME
0	XREF
0	CONTROLLED
0	CONTROL-TYPE
0	CONTROLLER

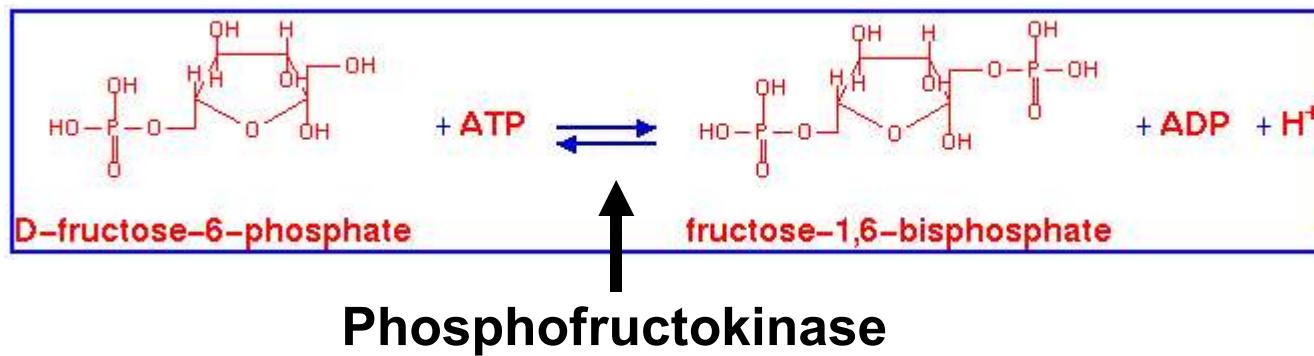
biochemicalReaction	
0	PARTICIPANTS
0	RIGHT
0	SYNONYMS
0	SHORT-NAME
0	SPONTANEOUS
0	COMMENT
0	AVAILABILITY
0	LEFT
0	DATA-SOURCE
0	NAME
0	XREF
0	DELTA-H
0	DELTA-S
0	EC-NUMBER
0	KEQ
0	DELTA-G

catalysis	
0	CONTROLLED
0	COMMENT
0	PARTICIPANTS
0	AVAILABILITY
0	CONTROL-TYPE
0	DATA-SOURCE
0	CONTROLLER
0	SYNONYMS
0	SHORT-NAME
0	NAME
0	XREF
0	DIRECTION
0	COFACTOR

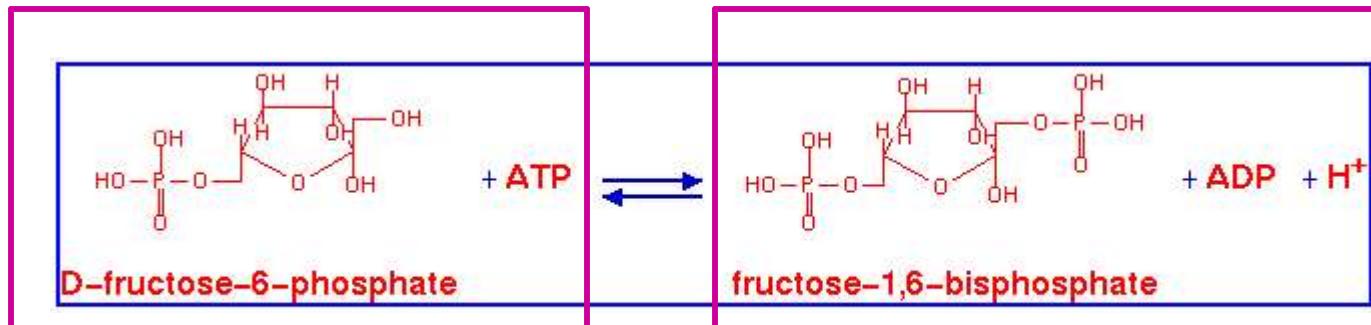
modulation	
0	CONTROLLED
0	COMMENT
0	PARTICIPANTS
0	AVAILABILITY
0	CONTROL-TYPE
0	DATA-SOURCE
0	CONTROLLER
0	SYNONYMS
0	SHORT-NAME
0	NAME
0	XREF

XML snippet

```
<bp:biochemicalReaction rdf:id="biochemicalReaction37">
  <bp:DATA-SOURCE rdf:resource="#dataSource14"/>
  <bp:LEFT>
    <bp:physicalEntityParticipant rdf:id="physicalEntityParticipant26">
      <bp:STOICHIOMETRIC-COEFFICIENT>1.0</bp:STOICHIOMETRIC-COEFFICIENT>
      <bp:PHYSICAL-ENTITY>
        <bp:smallMolecule rdf:id="smallMolecule27">
          <bp:SHORT-NAME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >a-D-glu-6-p</bp:SHORT-NAME>
          <bp:CHEMICAL-FORMULA rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >C6H13O9P</bp:CHEMICAL-FORMULA>
          <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >&lt;FONT FACE="Symbol">a&lt;/FONT>-D-glucose-6-phosphate</bp:SYNONYMS>
          <bp:XREF>
            <bp:unificationxref rdf:id="unificationxref30">
              <bp:ID rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
                >C00668</bp:ID>
              <bp:DB rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
                >KEGG</bp:DB>
            </bp:unificationxref>
          </bp:XREF>
        <bp:XREF rdf:resource="#unificationxref29"/>
        <bp:MOLECULAR-WEIGHT>260.14</bp:MOLECULAR-WEIGHT>
        <bp:AVAILABILITY rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >see http://www.amaze.ulb.ac.be/</bp:AVAILABILITY>
        <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >glucose-6-P</bp:SYNONYMS>
        <bp:DATA-SOURCE rdf:resource="#dataSource14"/>
        <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >alpha-D-glucose-6-p</bp:SYNONYMS>
        <bp:STRUCTURE>
          <bp:chemicalstructure rdf:id="chemicalstructure28">
            <bp:STRUCTURE-FORMAT>SMILES</bp:STRUCTURE-FORMAT>
            <bp:STRUCTURE-DATA>C1OP(=O)(O)O[C@H]1C(O)[C@H](O)[C@H](O)[C@H](O)O</bp:STRUCTURE-DATA>
          </bp:chemicalstructure>
        </bp:STRUCTURE>
        <bp:NAME>alpha-D-glucose 6-phosphate</bp:NAME>
        <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >alpha-D-glucose-6-phosphate</bp:SYNONYMS>
        <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          > D-glucose-6-P</bp:SYNONYMS>
        <bp:DATA-SOURCE rdf:resource="#KB_439584_Individual_47"/>
        </bp:smallMolecule>
      </bp:PHYSICAL-ENTITY>
      <bp:CELLULAR-LOCATION rdf:resource="#openControlledVocabulary15"/>
    </bp:physicalEntityParticipant>
  </bp:LEFT>
  <bp:DELTA-G rdf:datatype="http://www.w3.org/2001/XMLSchema#double"
    >0.4</bp:DELTA-G>
  <bp:SYNONYMS rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >alpha-D-Glucose 6-phosphate &lt;=> beta-D-Fructose 6-phosphate </bp:SYNONYMS>
  <bp:RIGHT>
    <bp:physicalEntityParticipant rdf:id="physicalEntityParticipant38">
      <bp:CELLULAR-LOCATION rdf:resource="#openControlledVocabulary15"/>
      <bp:PHYSICAL-ENTITY>
        <bp:smallMolecule rdf:id="smallMolecule39">
```



Biochemical Reaction Glycolysis Pathway

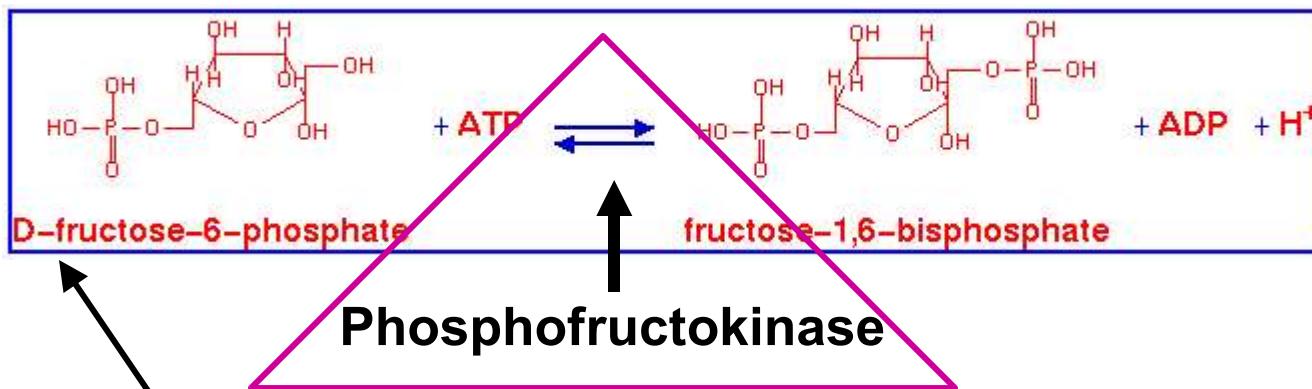


Left

Right

biochemicalReaction	
①	PARTICIPANTS
①	RIGHT
①	SYNONYMS
①	SHORT-NAME
①	SPONTANEOUS
①	COMMENT
①	AVAILABILITY
①	LEFT
①	DATA-SOURCE
①	NAME
①	XREF
②	DELTA-H
②	DELTA-S
②	EC-NUMBER
②	KEQ
②	DELTA-G

EC # 2.7.1.11

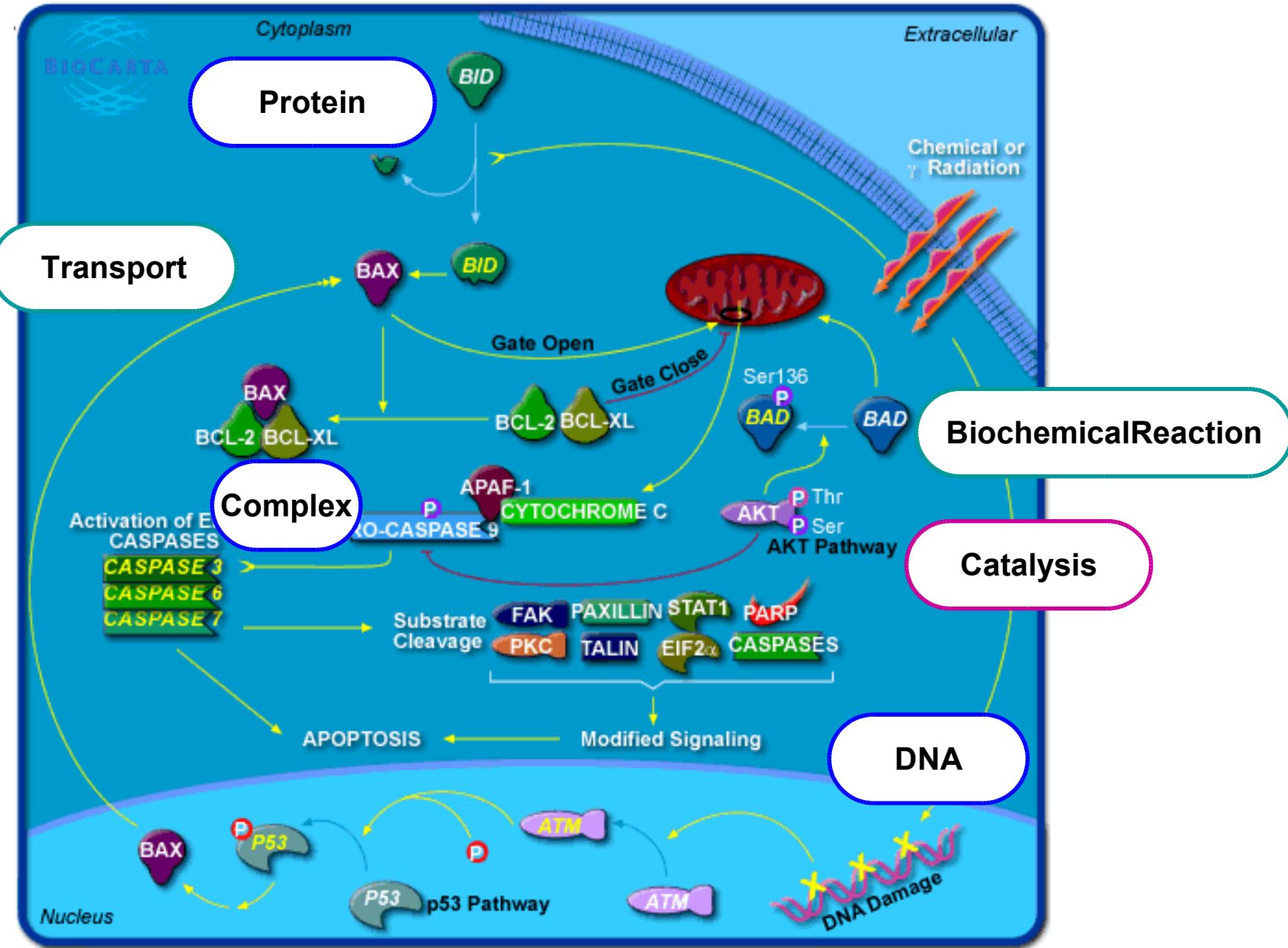


Controlled

Controller

catalysis	
0	CONTROLLED
0	COMMENT
0	PARTICIPANTS
0	AVAILABILITY
0	CONTROL-TYPE
0	DATA-SOURCE
0	CONTROLLER
0	SYNONYMS
0	SHORT-NAME
0	NAME
0	XREF
0	DIRECTION
0	COFACTOR

Direction: reversible



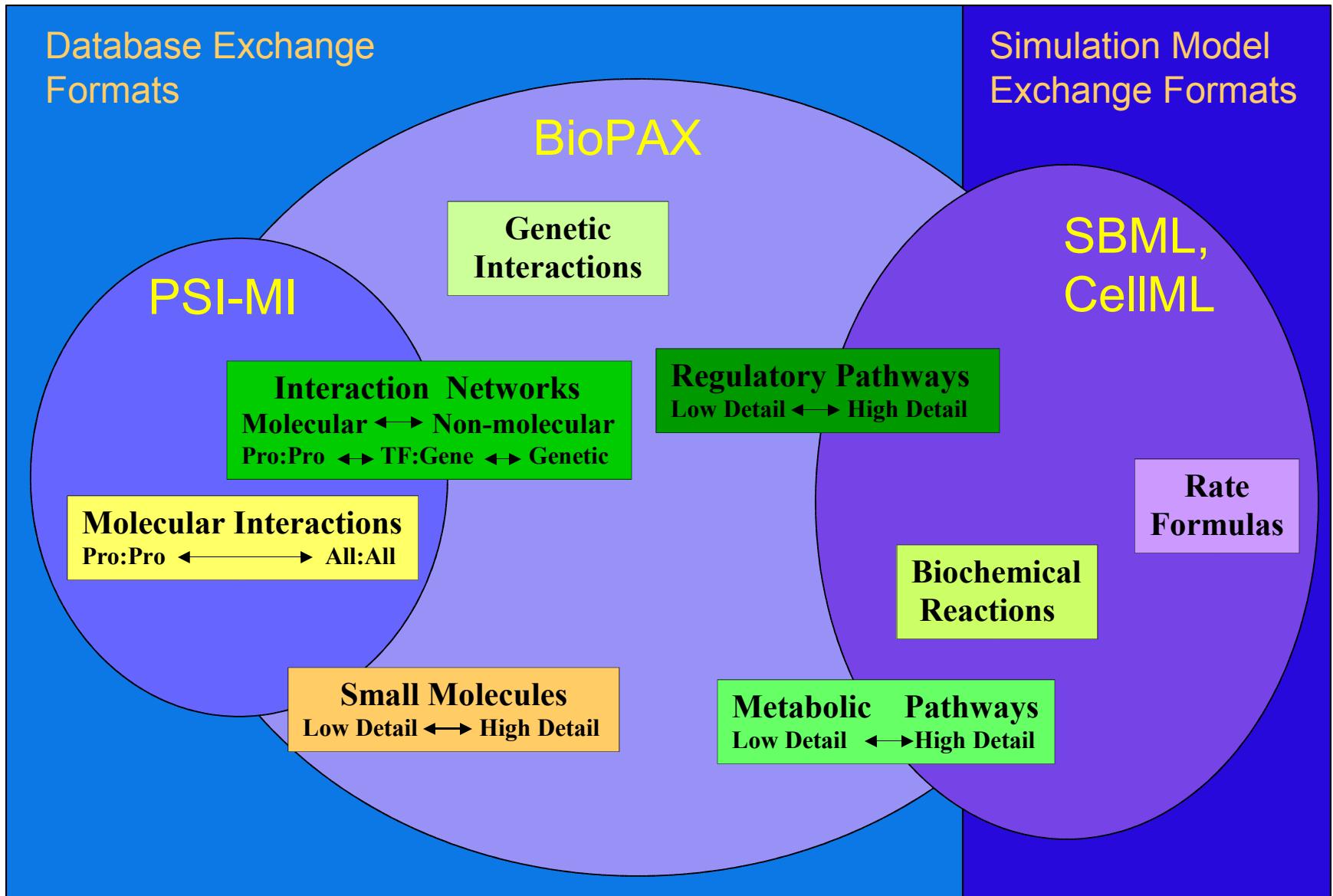
Controlled Vocabularies (CVs)

- 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/>

Worked examples

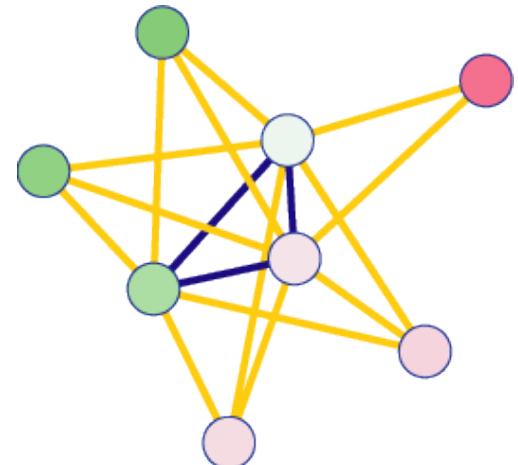
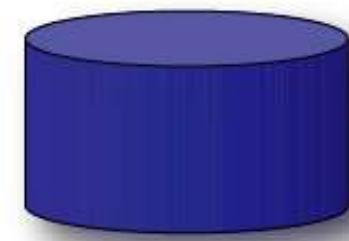
- Metabolic pathway
 - EcoCyc Glycolysis (energy metabolism pathway)
- Protein-protein interaction
 - Proteomics, PSI-MI
- Signaling pathway step
 - Reactome CHK2-ATM
- Switch to Protégé
- Available from biopax.org
 - <http://www.biopax.org/Downloads/Level2v1.0/biopax-level2.zip>

Exchange Formats in the Pathway Data Space

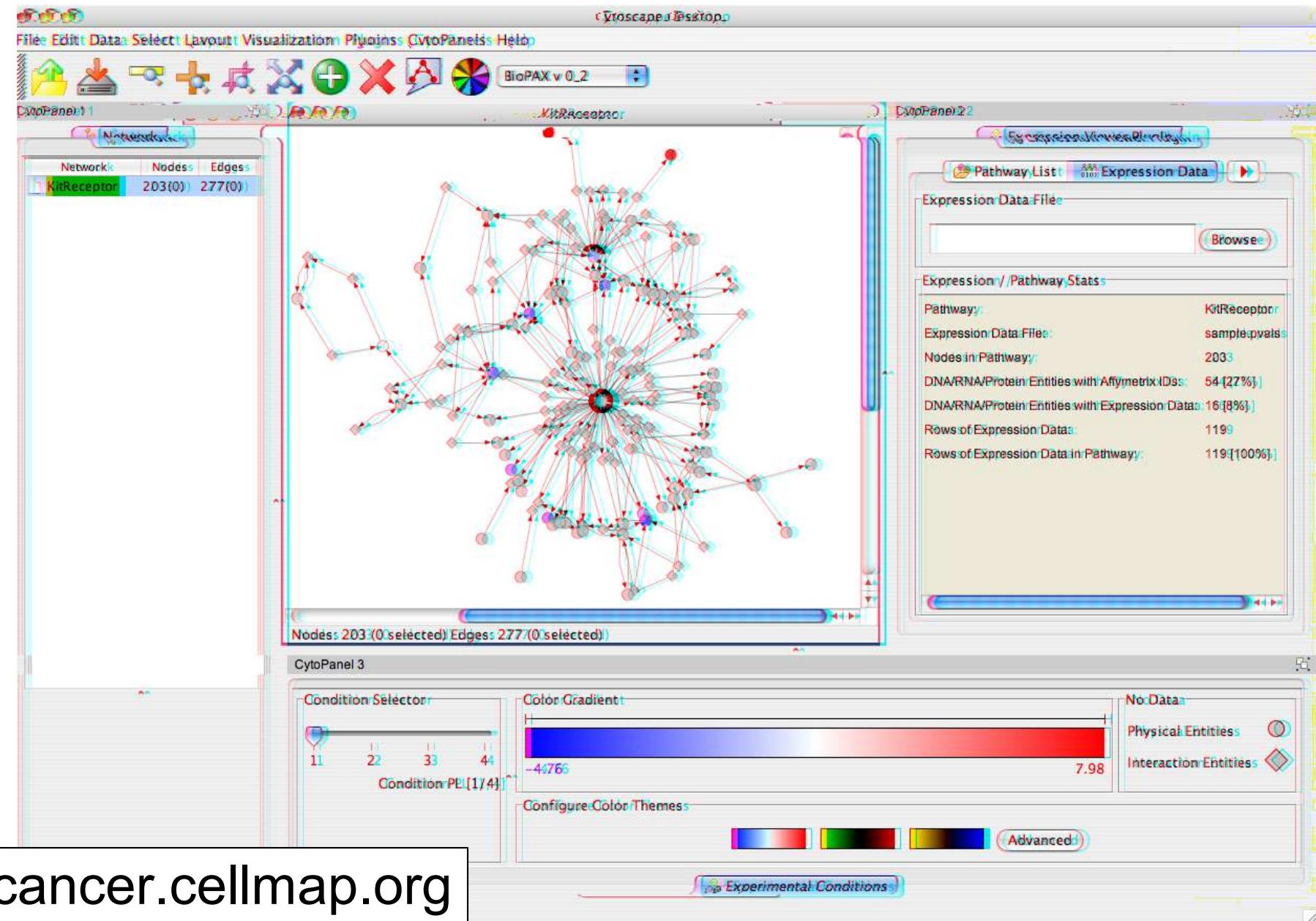


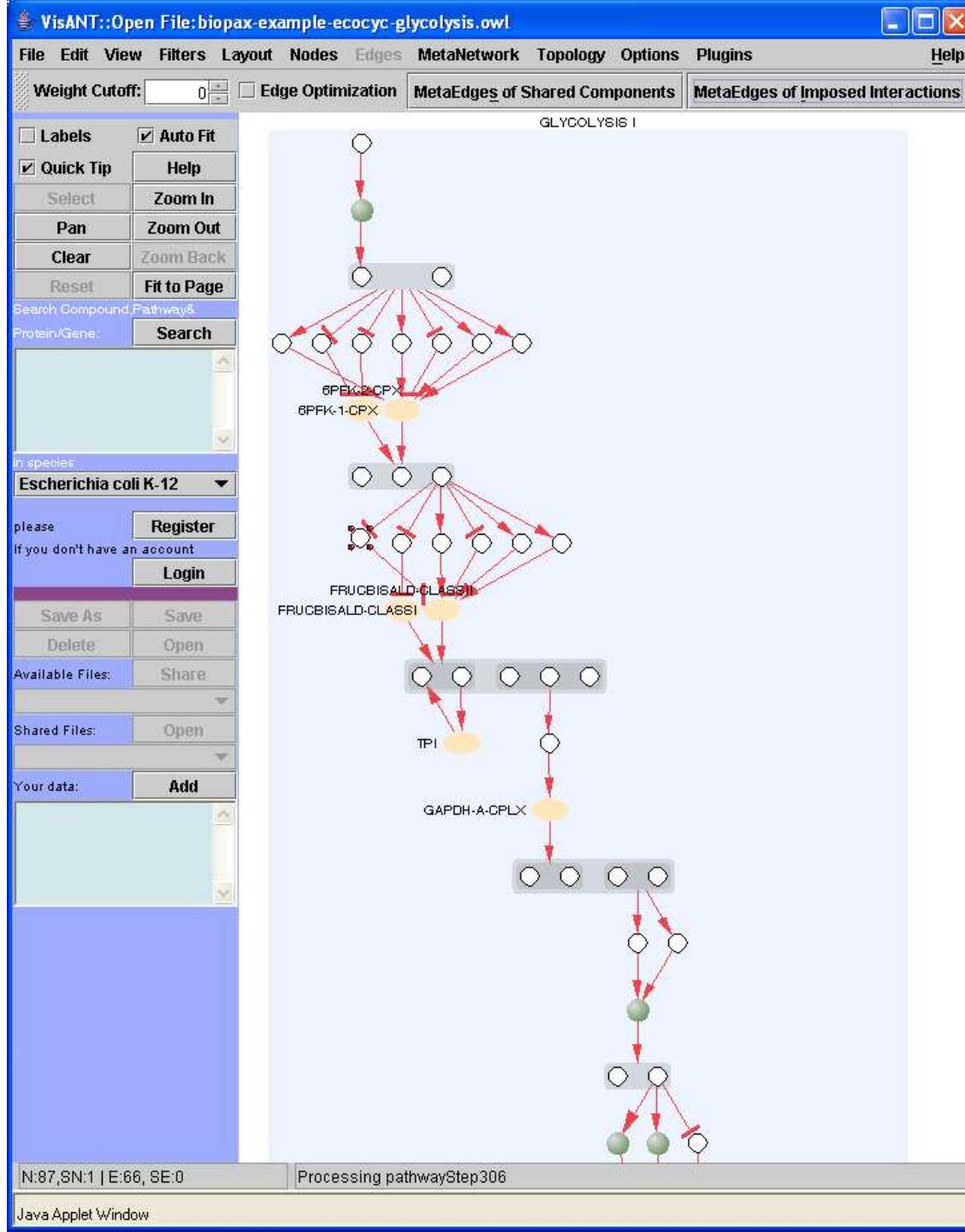
Using BioPAX

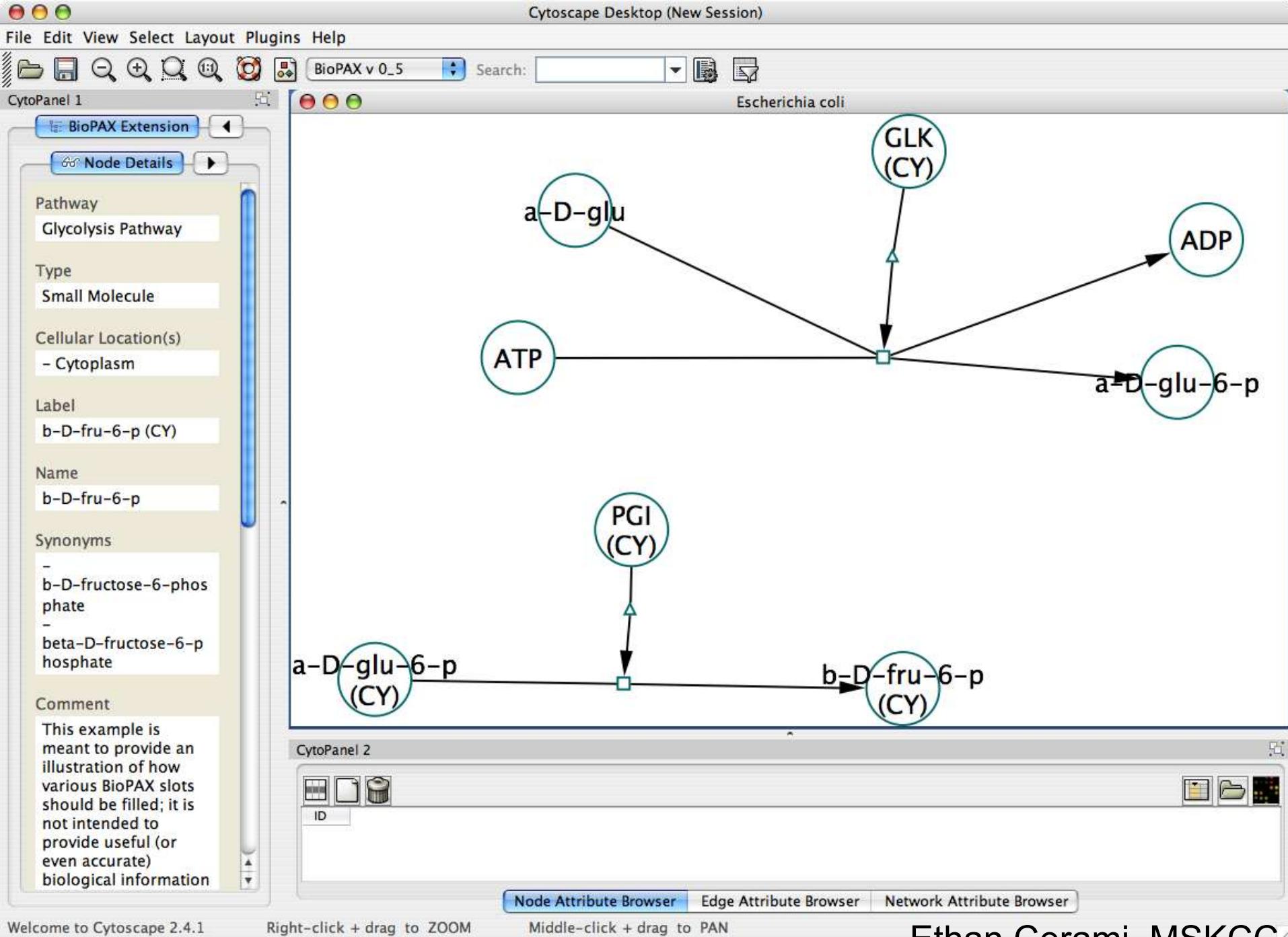
- Databases
 - BioCyc (EcoCyc, MetaCyc, many pathway genome databases)
 - KEGG (available soon – KEGG, aMAZE, Sander)
 - MSKCC Cancer Pathway Resource
 - Reactome
 - PSI-MI (via converter)
 - Switch to Pathguide
- Tools
 - cPath, Cytoscape, GenMAPP, PATIKA, QPACA, VisANT
- caBIG



The Cancer Cell Map





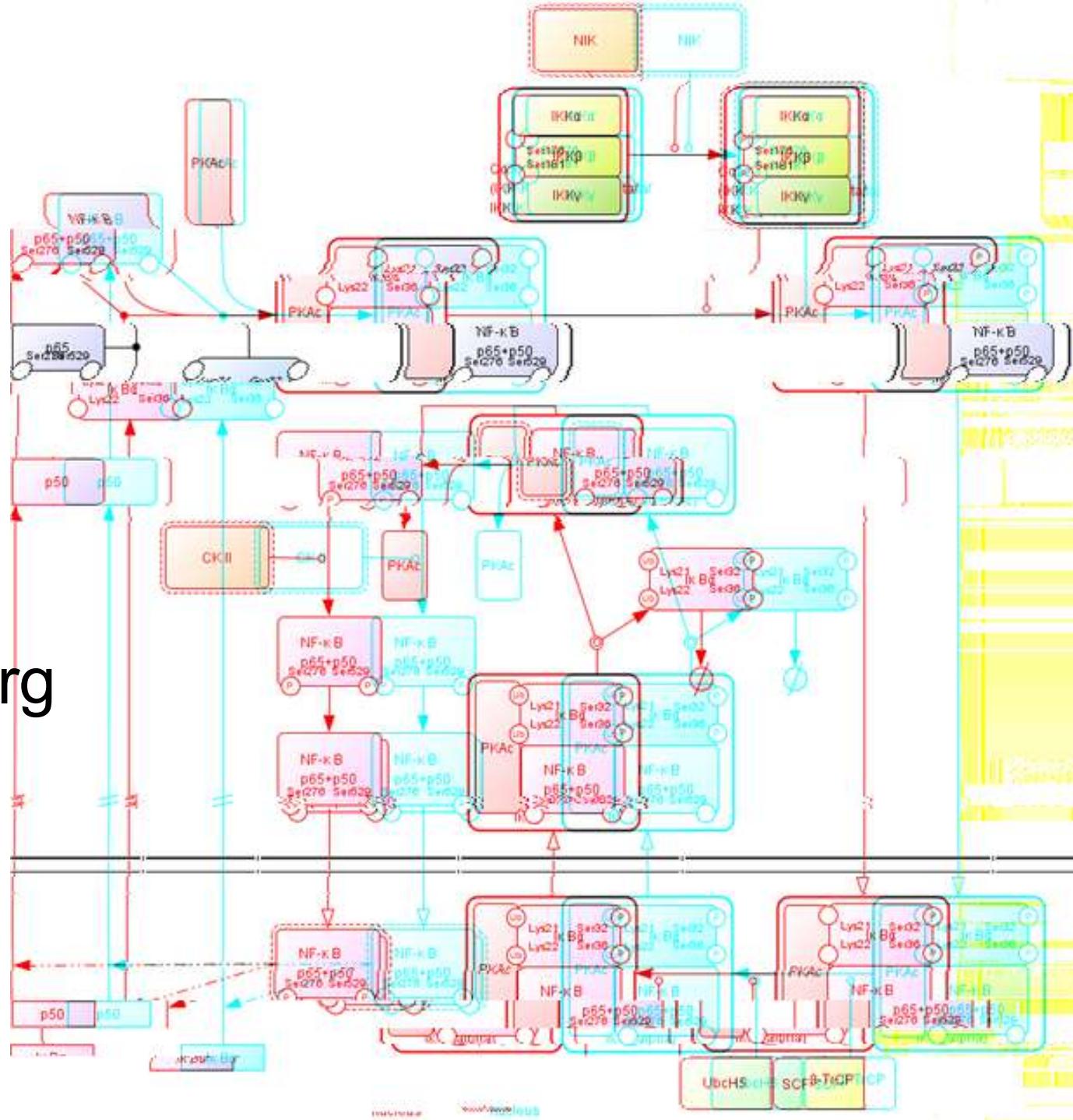


Switch to Cytoscape

- Load BioPAX pathway from Reactome (reactome.org)
 - http://reactome.org/cgi-bin/biopaxexporter?DB=gk_current&ID=195721
- Load, view + lay out
- Extract UniProt IDs from Cytoscape attributes

Systems Biology Graphical Notation

<http://sbgn.org>
In progress



Software Development

- PaxTools
 - Open source Java
 - Read/write BioPAX files (Level 1,2)
 - Object model in memory that can be populated and queried
 - Validation on create, read (under development by MSKCC, OHSU)
 - <http://biopax.cvs.sourceforge.net/biopax/Paxerve/>

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

How to participate and contribute

- Visit biopax.org and join the discussion mailing list
 - biopax-discuss@biopax.org
- Make pathway data available in BioPAX
- Build software that supports BioPAX
- Contribute BioPAX worked examples, documentation and specification reviews
- Spread the word about BioPAX

BioPAX Supporting Groups

Current Participants

- Memorial Sloan-Kettering Cancer Center: E.Demir, M. Cary, C. Sander
- University of Toronto: G. Bader
- SRI Bioinformatics Research Group: P. Karp, S. Paley, J. Pick
- Bilkent University: U. Dogrusoz
- Université Libre de Bruxelles: C. Lemmerling
- CBRC Japan: K. Fukuda
- Dana Farber Cancer Institute: J. Zucker
- Millennium: J. Rees, A. Ruttenberg
- Cold Spring Harbor/EBI: G. Wu, M. Gillespie, P. D'Eustachio, I. Vastrik, L. Stein
- BioPathways Consortium: J. Luciano, E. Neumann, A. Regev, V. Schachter
- Argonne National Laboratory: N. Maltsev, E. Marland, M. Syed
- Harvard: F. Gibbons
- AstraZeneca: E. Pichler
- BIOBASE: E. Wingender, F. Schacherer
- NCI: M. Aladjem, C. Schaefer
- Università di Milano Bicocca, Pasteur, Rennes: A. Splendiani
- Vassar College: K. Dahlquist
- Columbia: A. Rzhetsky

Collaborating Organizations

- Proteomics Standards Initiative (PSI)
- Systems Biology Markup Language (SBML)
- CellML
- Chemical Markup Language (CML)

Databases

- BioCyc, WIT, KEGG, BIND, PharmGKB, aMAZE, INOH, Transpath, Reactome, PATIKA, eMIM, NCI PID, CellMap

Wouldn't be possible without

Gene Ontology

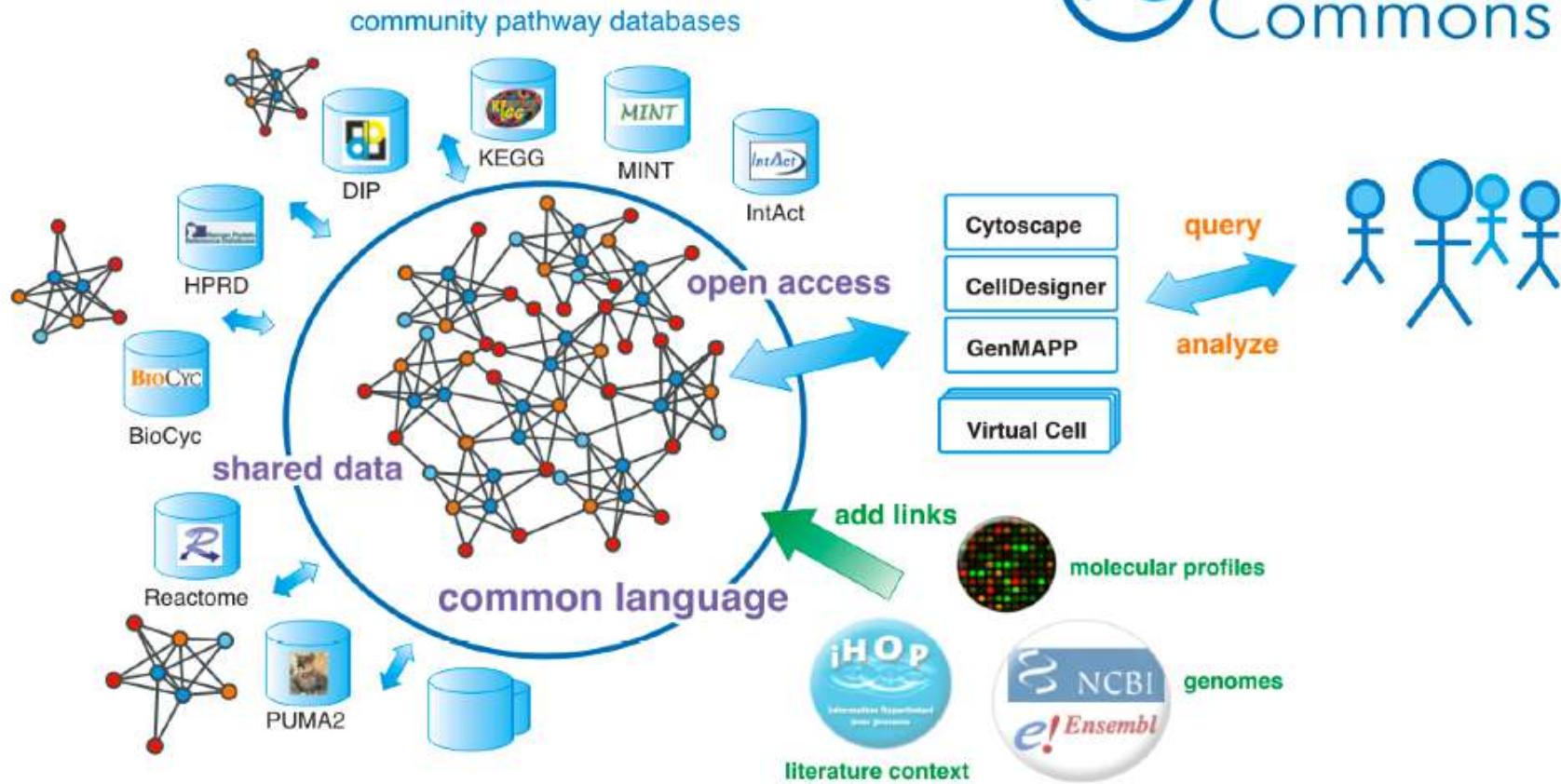
Protégé, U.Manchester, Stanford

Grants/Support

- Department of Energy (Workshop)
-  caBIG

Aim: Convenient Access to Pathway Information

<http://www.pathwaycommons.org>



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