http://www.pathwaycommons.org

Pathway Commons

A public library of biological pathways

July.14 2008 - Bioinformatics to Systems Biology 2008

Gary Bader
University of Toronto

http://baderlab.org

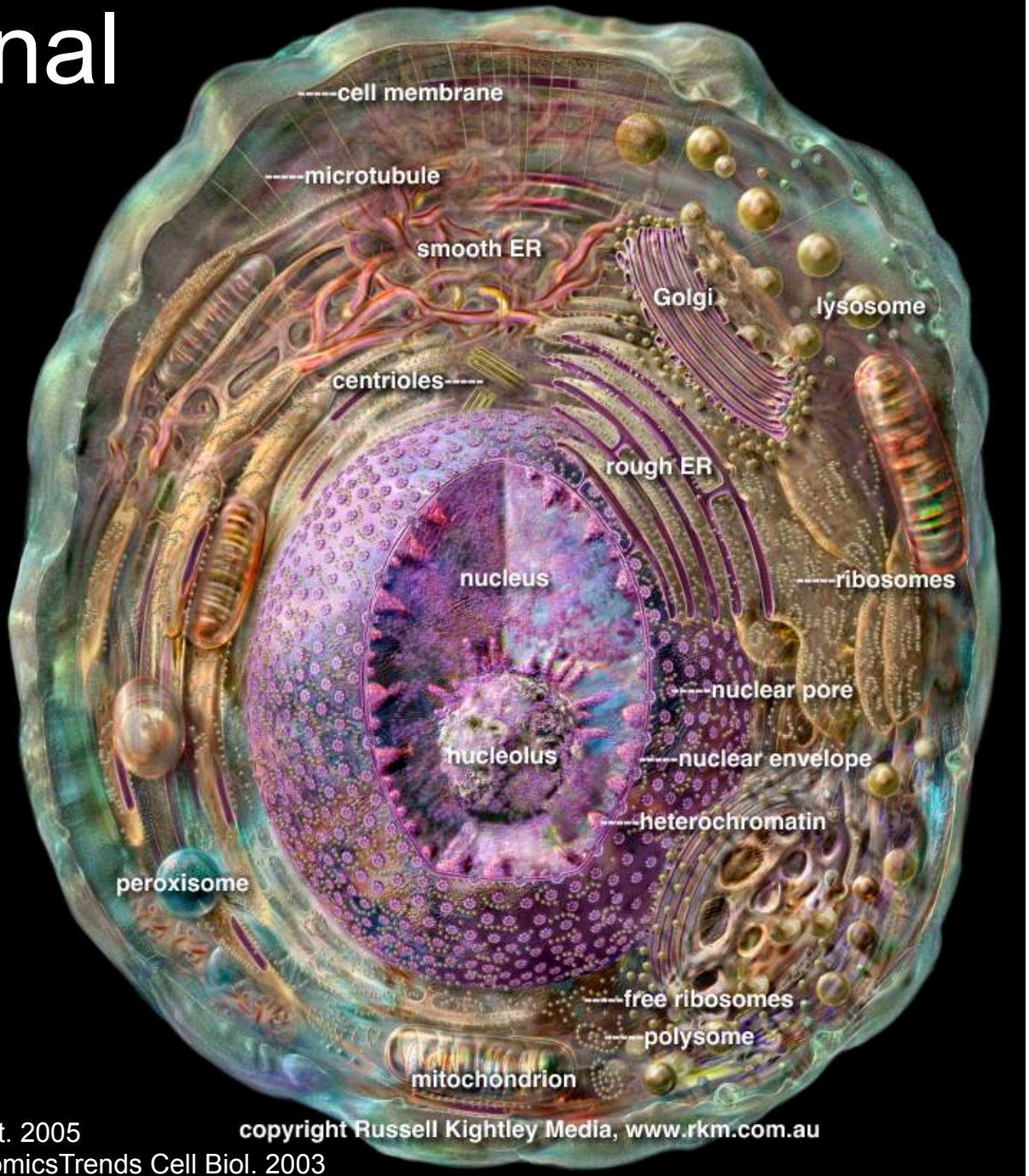
Chris Sander
MSKCC, New York

http://cbio.mskcc.org

Computational Cell Map

Map the cell
- Predict map from genome
- Multiple perturbation mapping
- Active cell map
- Map visualization and analysis software

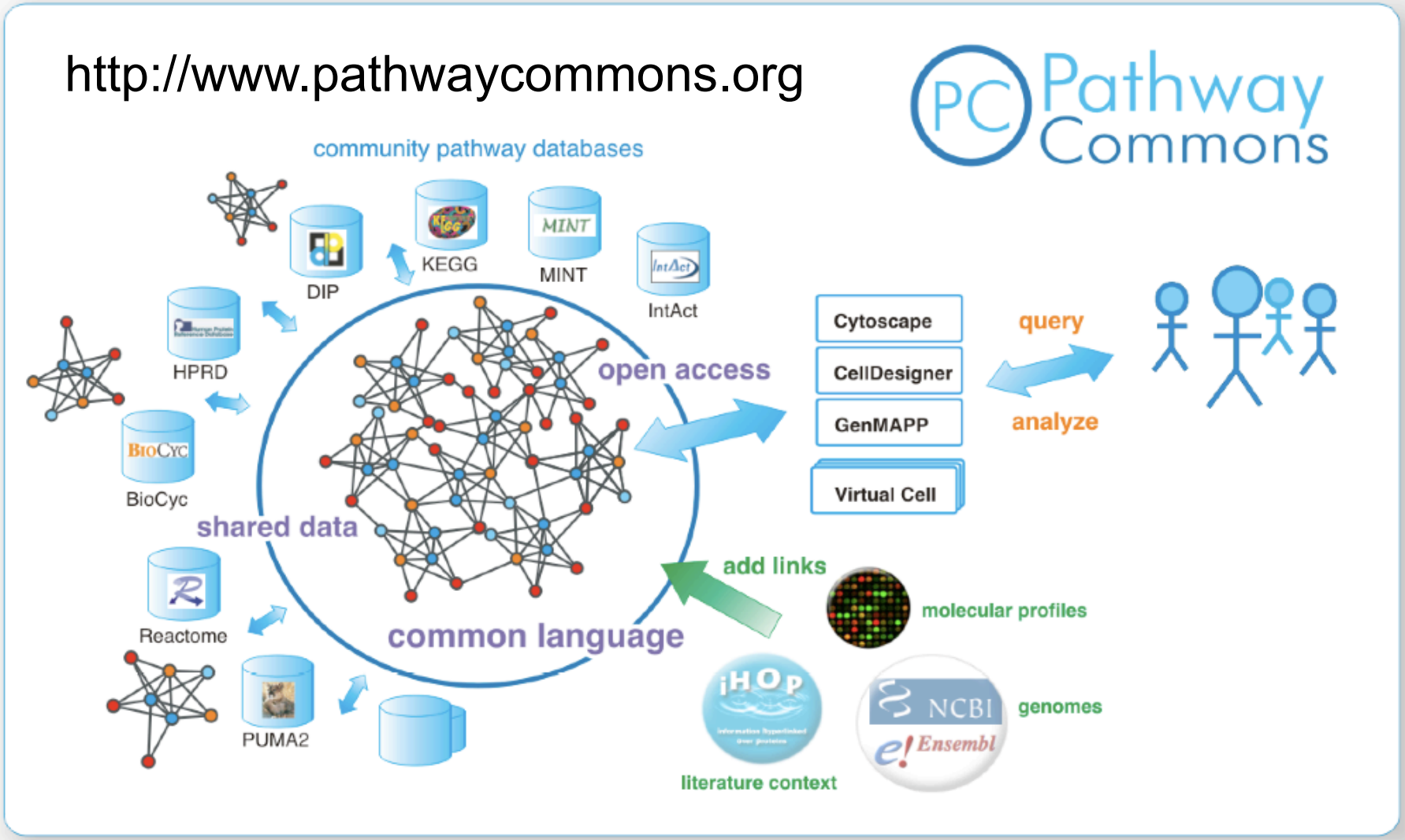
Read map to understand
- Cell processes
- Gene function
- Disease effects
- Map evolution



Cary MP et al. Pathway information… FEBS Lett. 2005
Bader GD et al. Functional genomics and proteomicsTrends Cell Biol. 2003

Aim: Convenient Access to Pathway Information

http://www.pathwaycommons.org

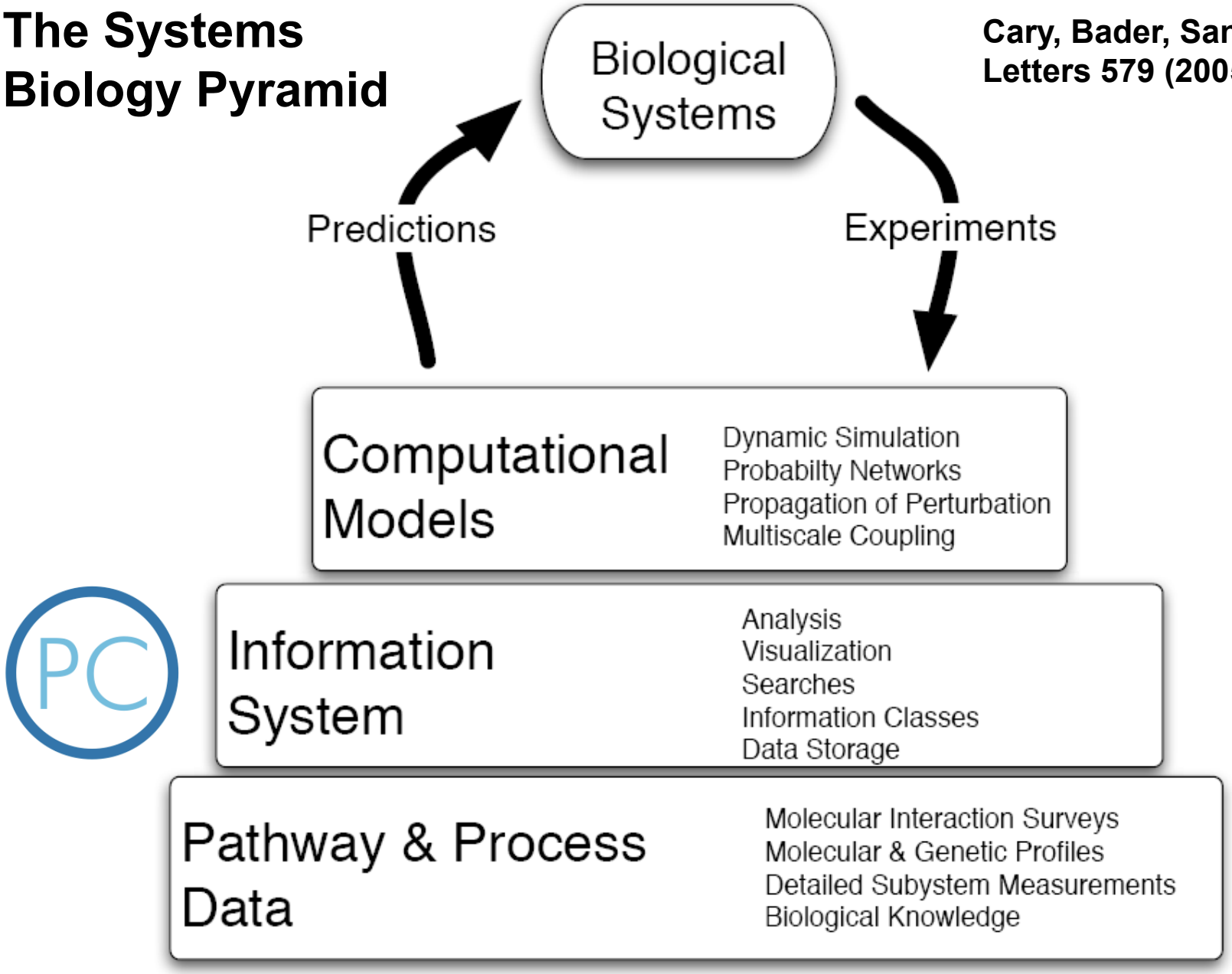


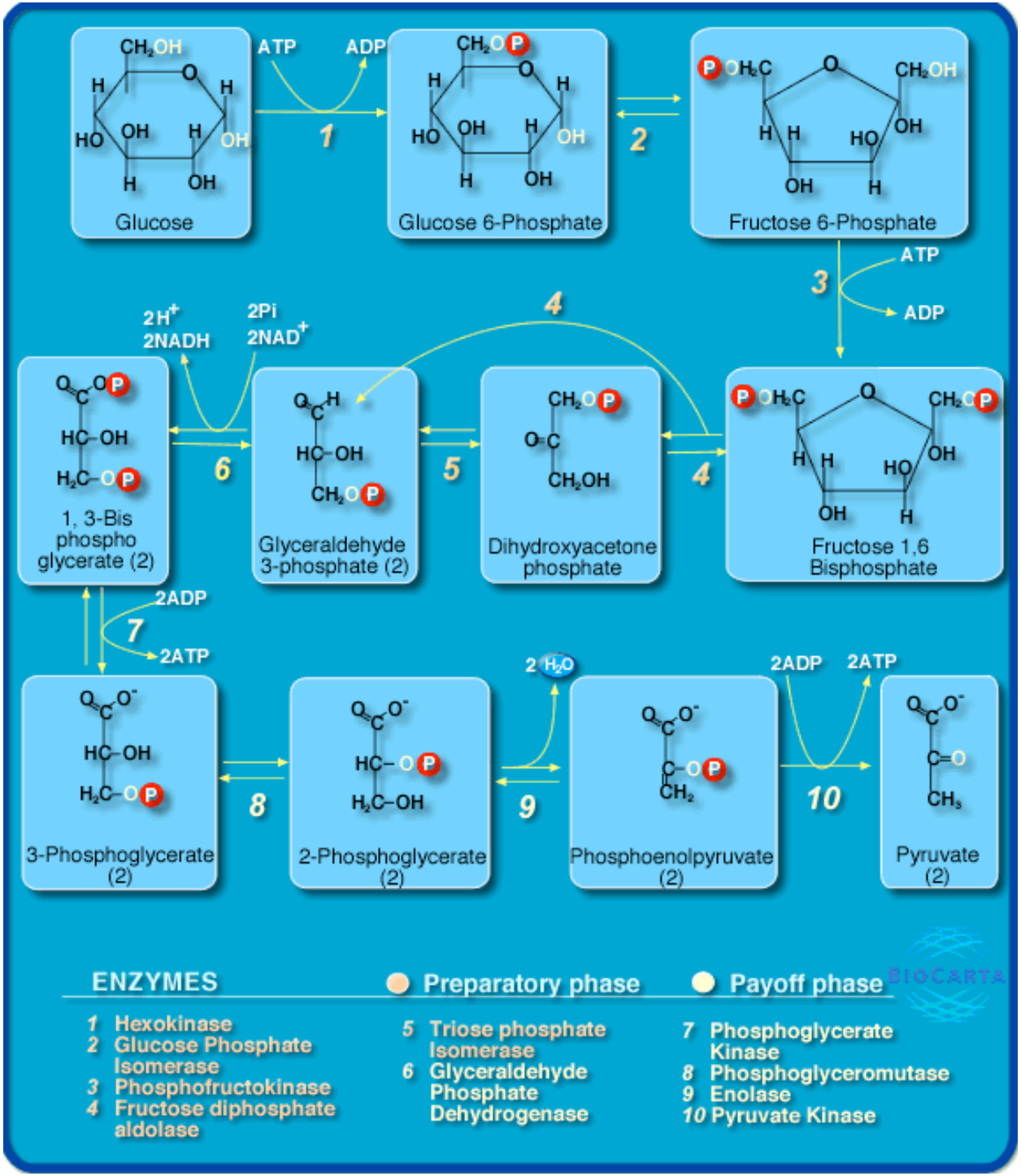
Facilitate creation and communication of pathway data
Aggregate pathway data in the public domain
Provide easy access for pathway analysis

Long term: Converge to integrated cell map

The Systems Biology Pyramid

Cary, Bader, Sander, FEBS Letters 579 (2005) 1815-20




Glucose
ATP
ADP
1
Glucose 6-Phosphate
2
Fructose 6-Phosphate
3
ATP
ADP
4
2H+
2NADH
2Pi
2NAD+
1, 3-Bis phospho glycerate (2)
6
Glyceraldehyde 3-phosphate (2)
5
Dihydroxyacetone phosphate
4
Fructose 1,6 Bisphosphate
2ADP
7
2ATP
3-Phosphoglycerate (2)
8
2-Phosphoglycerate (2)
2 H2O
9
Phosphoenolpyruvate (2)
2ADP
2ATP
10
Pyruvate (2)
ENZYMES
Preparatory phase
Payoff phase
1 Hexokinase
2 Glucose Phosphate Isomerase
3 Phosphofructokinase
4 Fructose diphosphate aldolase
5 Triose phosphate Isomerase
6 Glyceraldehyde Phosphate Dehydrogenase
7 Phosphoglycerate Kinase
8 Phosphoglyceromutase
9 Enolase
10 Pyruvate Kinase

Signaling Pathway

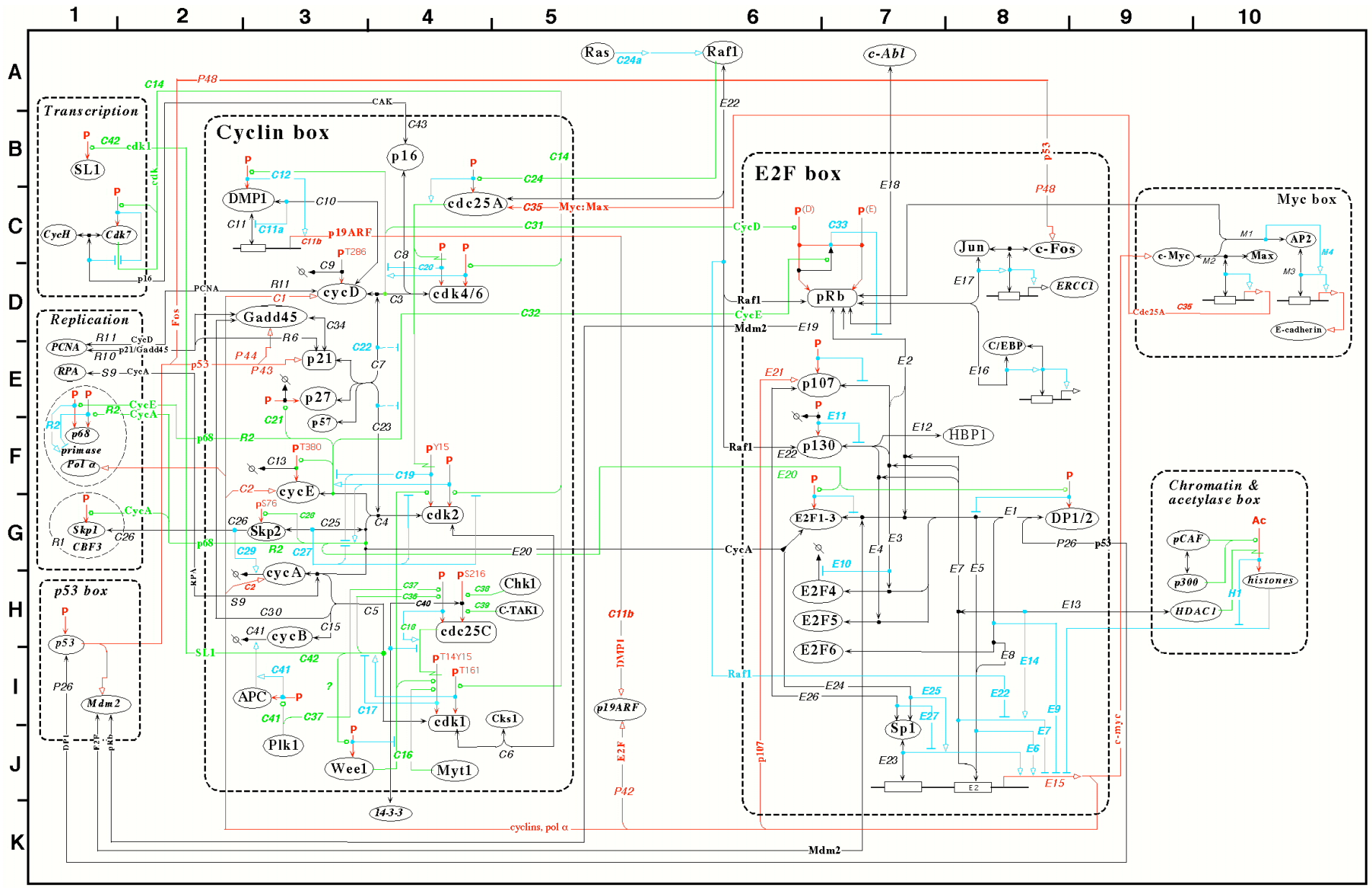
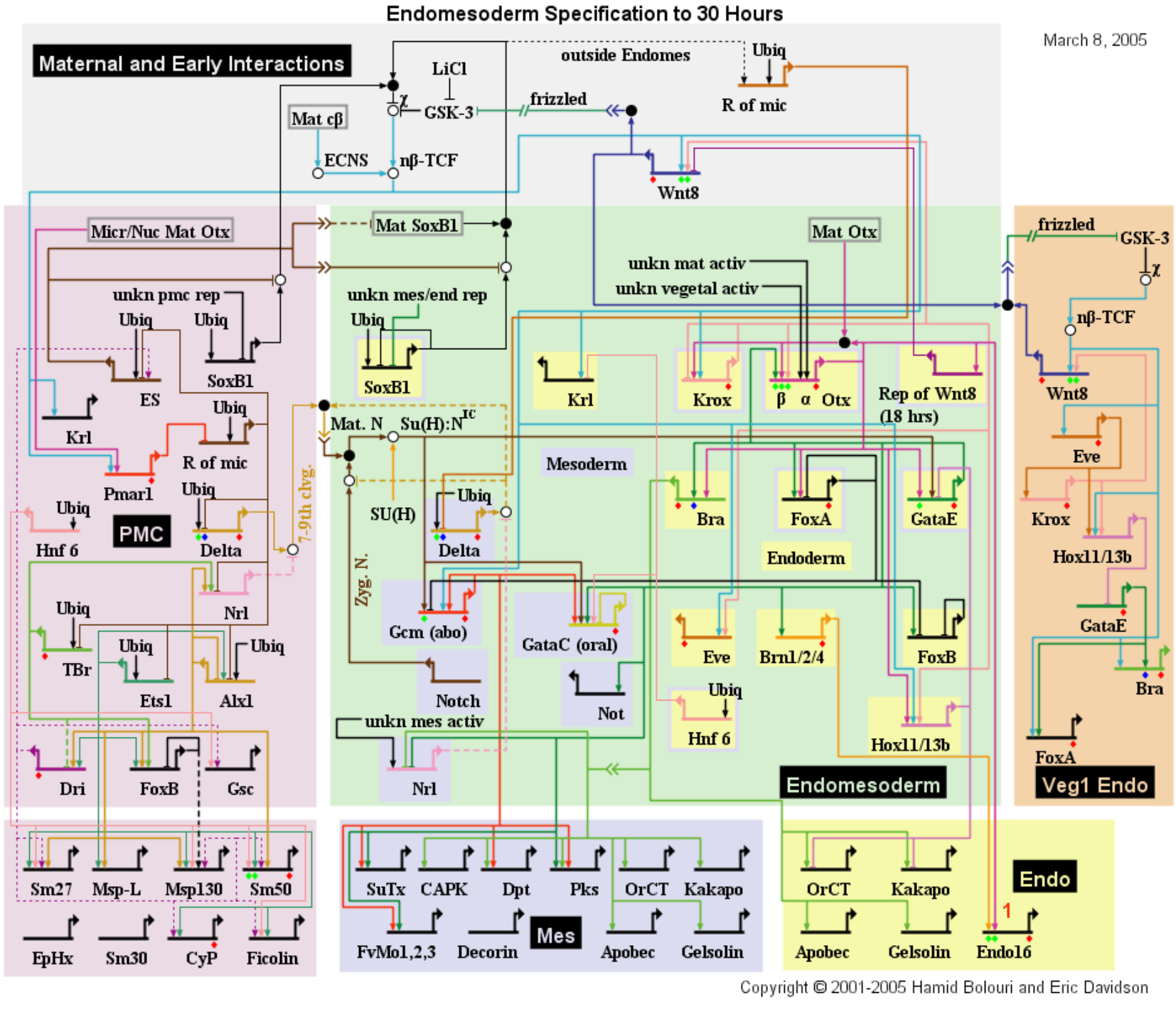
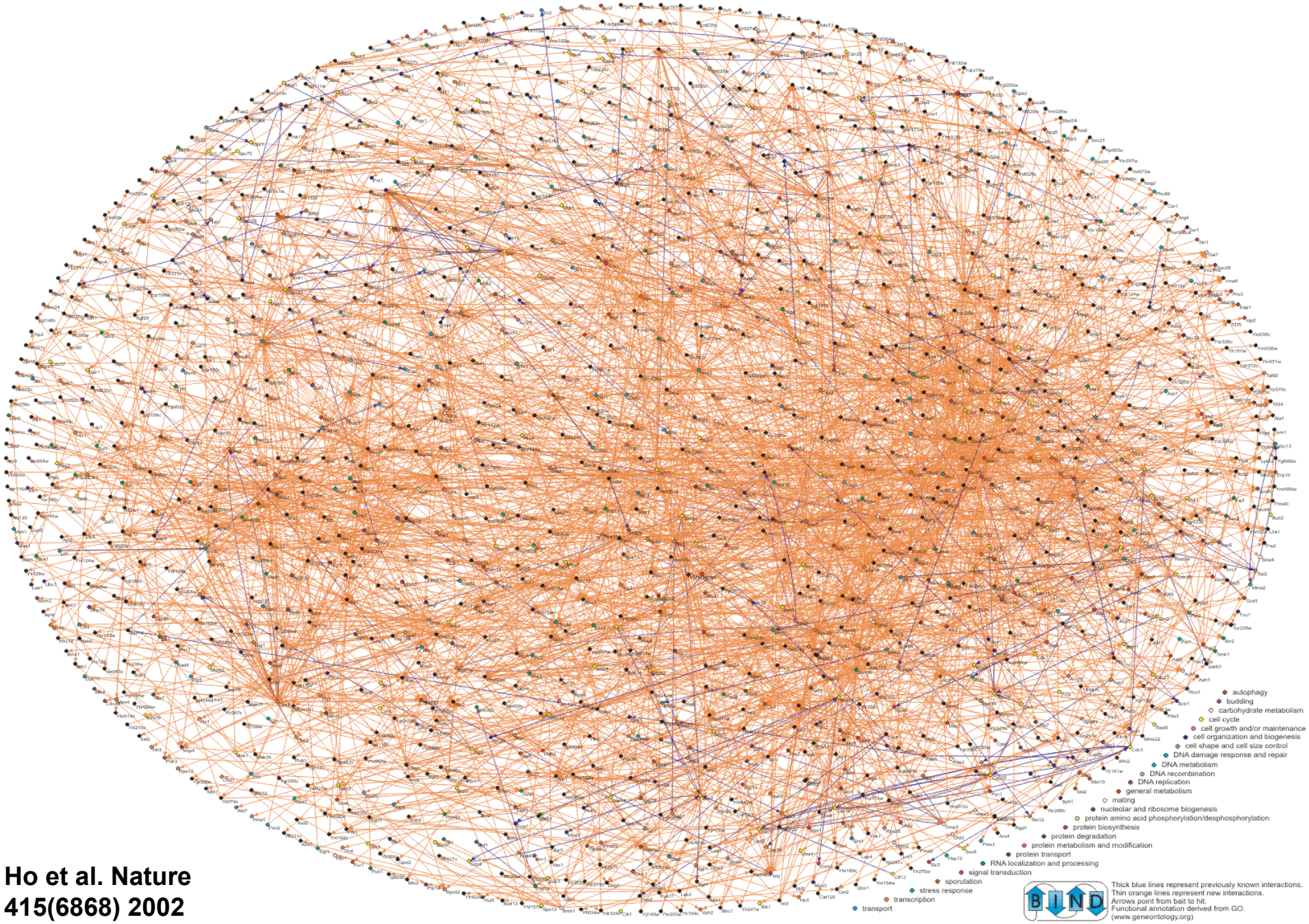


Figure 6A: The Cyclin - E2F cell cycle control system (version 3a - June 8, 1999)

http://discover.nci.nih.gov/kohnk/interaction_maps.html


Endomesoderm Specification to 30 Hours
March 8, 2005
Maternal and Early Interactions
LiCl
GSK-3
Mat cβ
ECNS
nβ-TCF
outside Endomes
Ubiq
R of mic
frizzled
Wnt8
Micr/Nuc Mat Otx
Mat SoxB1
Mat Otx
unkn mat activ
unkn vegetal activ
unkn pmc rep
unkn mes/end rep
SoxB1
ES
Krl
Pmar1
R of mic
PMC
Hnf 6
Delta
Nrl
TBr
Ets1
Alx1
Dri
FoxB
Gsc
7-9th clvg.
Mat. N
Su(H):N^IC
SU(H)
Zyg. N.
Gcm (abo)
GataC (oral)
Notch
Not
unkn mes activ
Nrl
Krox
β α Otx
Rep of Wnt8
(18 hrs)
Mesoderm
Bra
FoxA
GataE
Endoderm
Eve
Brn1/2/4
FoxB
Hnf 6
Hox11/13b
Endomesoderm
Veg1 Endo
Sm27
Msp-L
Msp130
Sm50
EpHx
Sm30
CyP
Ficolin
SuTx
CAPK
Dpt
Pks
OrCT
Kakapo
FvMo1,2,3
Decorin
Mes
Apobec
Gelsolin
Endo
Endo16
1
Copyright © 2001-2005 Hamid Bolouri and Eric Davidson

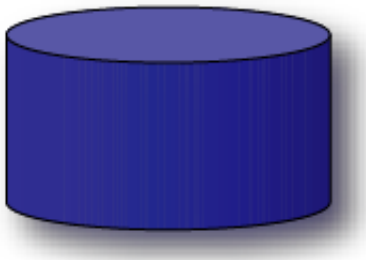
Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry

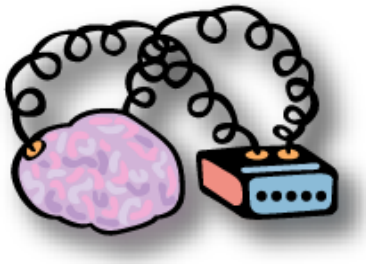


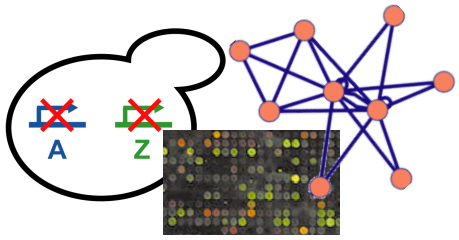
Ho et al. Nature
415(6868) 2002

Pathway Information

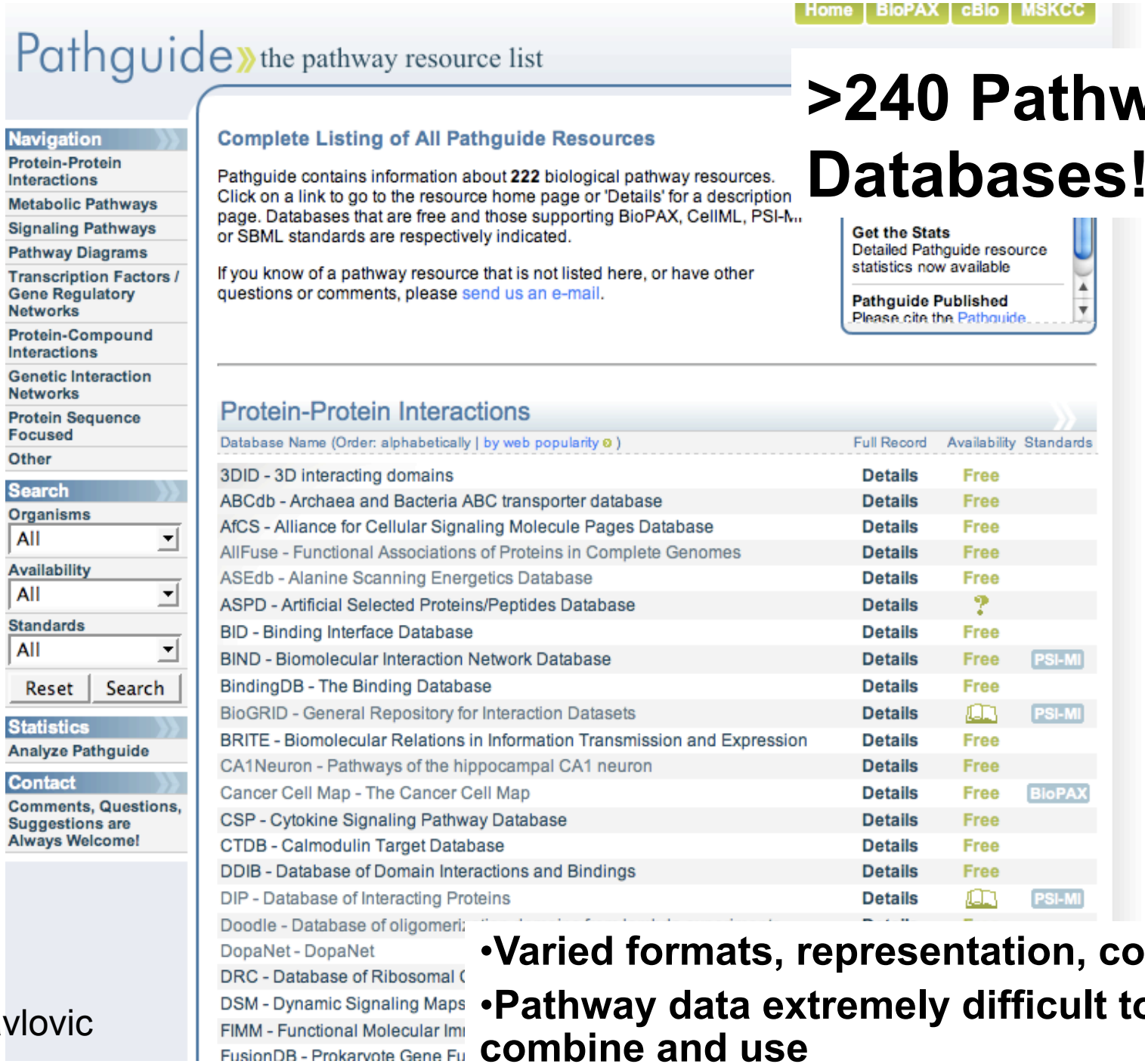
- Databases
  - Fully electronic
  - Easily computer readable
- Literature
  - Increasingly electronic
  - Human readable
- Biologist's brains
  - Richest data source
  - Limited bandwidth access
- Experiments
  - Basis for models







http://pathguide.org

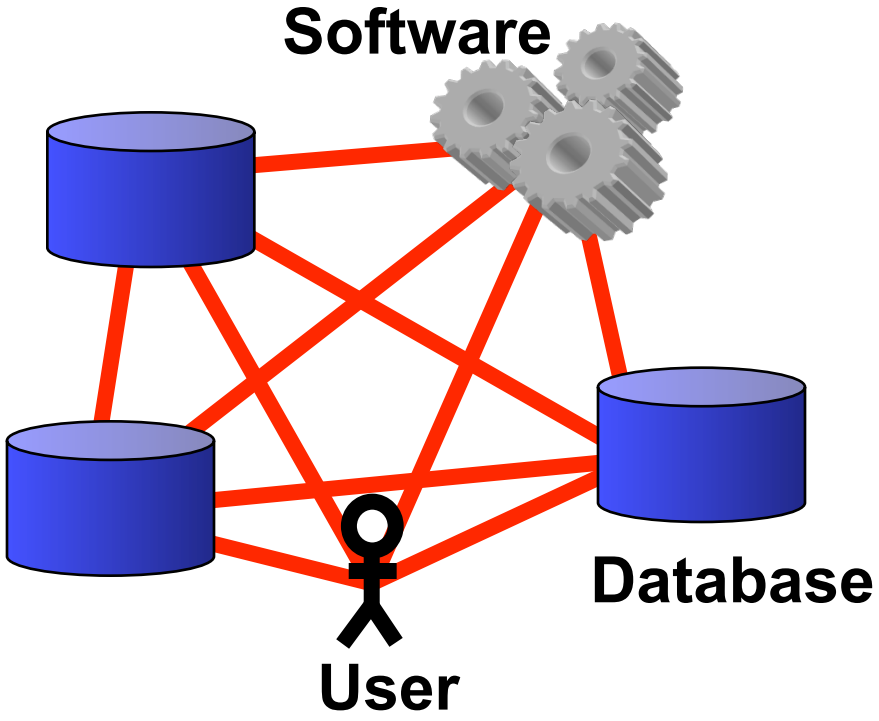


>240 Pathway Databases!

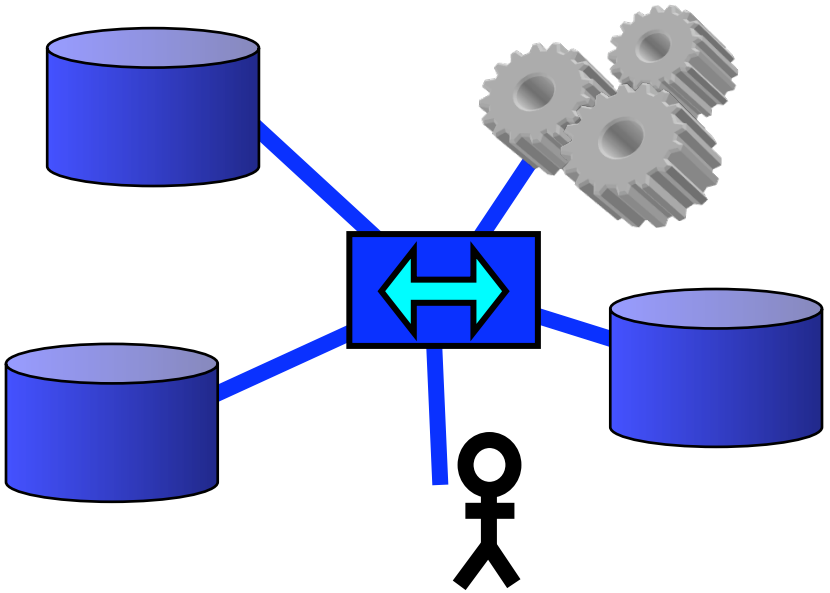
- **Varied formats, representation, coverage**
- **Pathway data extremely difficult to combine and use**

Vuk Pavlovic

Biological Pathway Exchange (BioPAX)



Before BioPAX
>100 DBs and tools
Tower of Babel



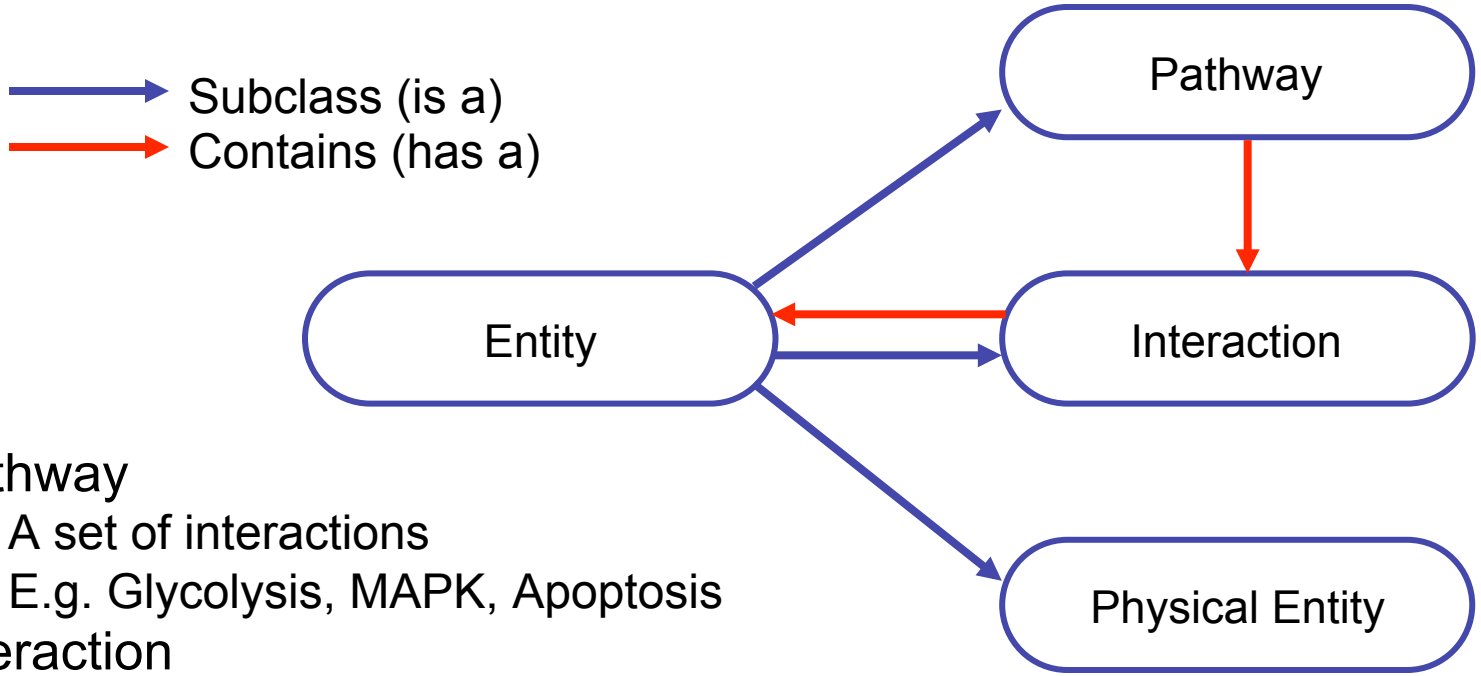
After BioPAX
Unifying language

Reduces work, promotes collaboration, increases accessibility

BioPAX Pathway Language

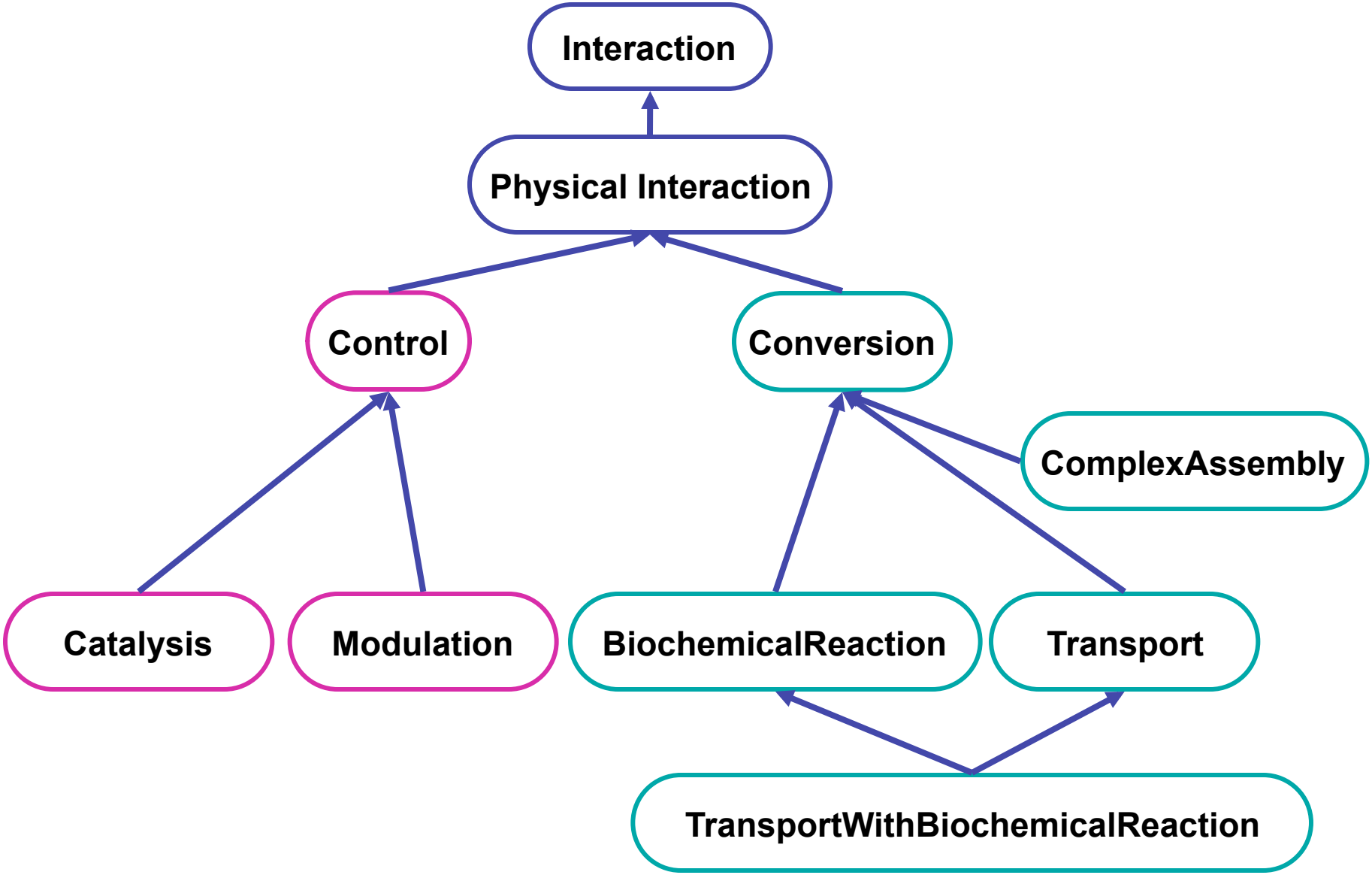
- Represent:
  - Metabolic pathways
  - Signaling pathways
  - Protein-protein, molecular interactions
  - Gene regulatory pathways
  - Genetic interactions
- Community effort: pathway databases distribute pathway information in standard format

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 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. Lemer
- 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

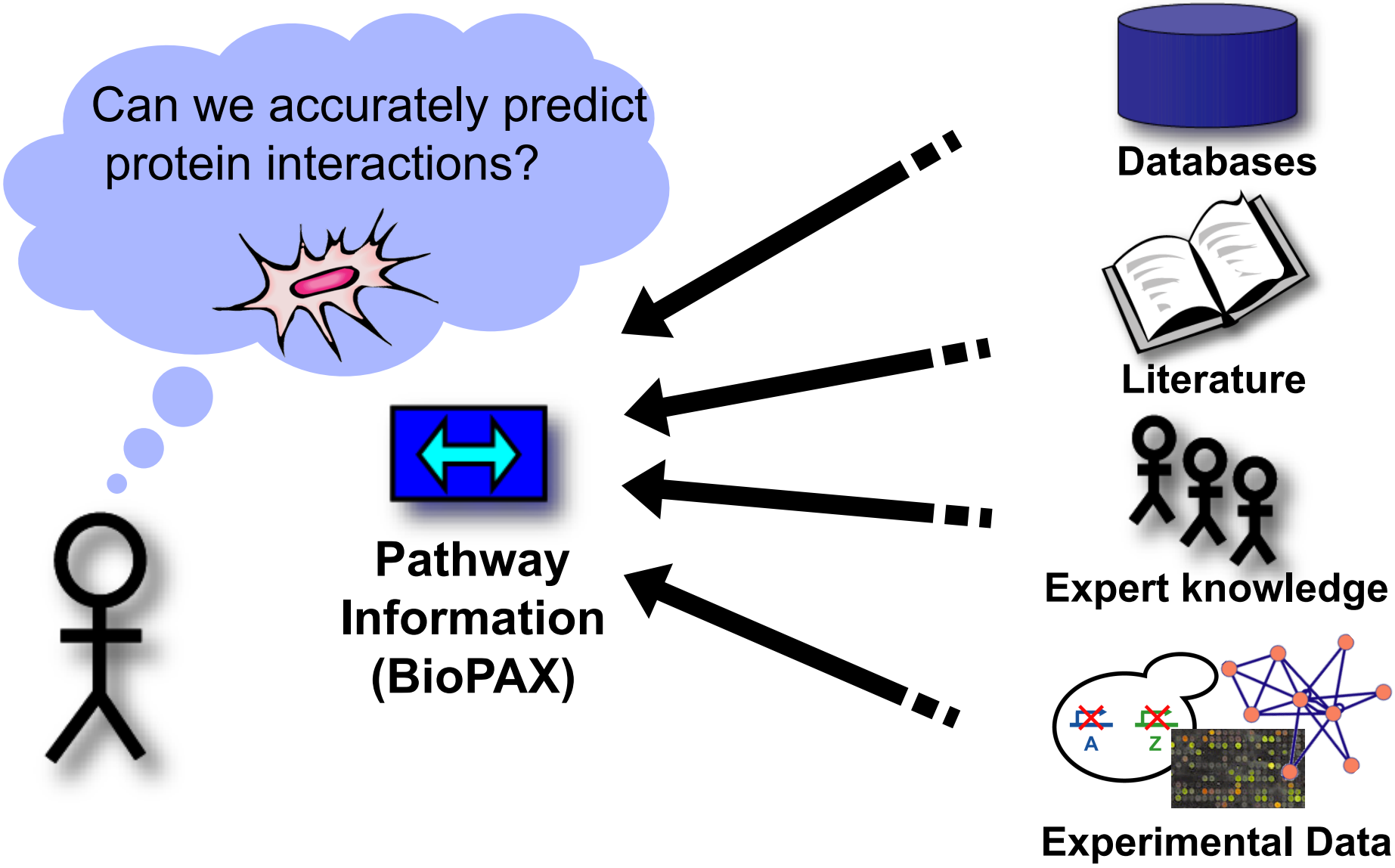
Gene Ontology

Protégé, U.Manchester, Stanford

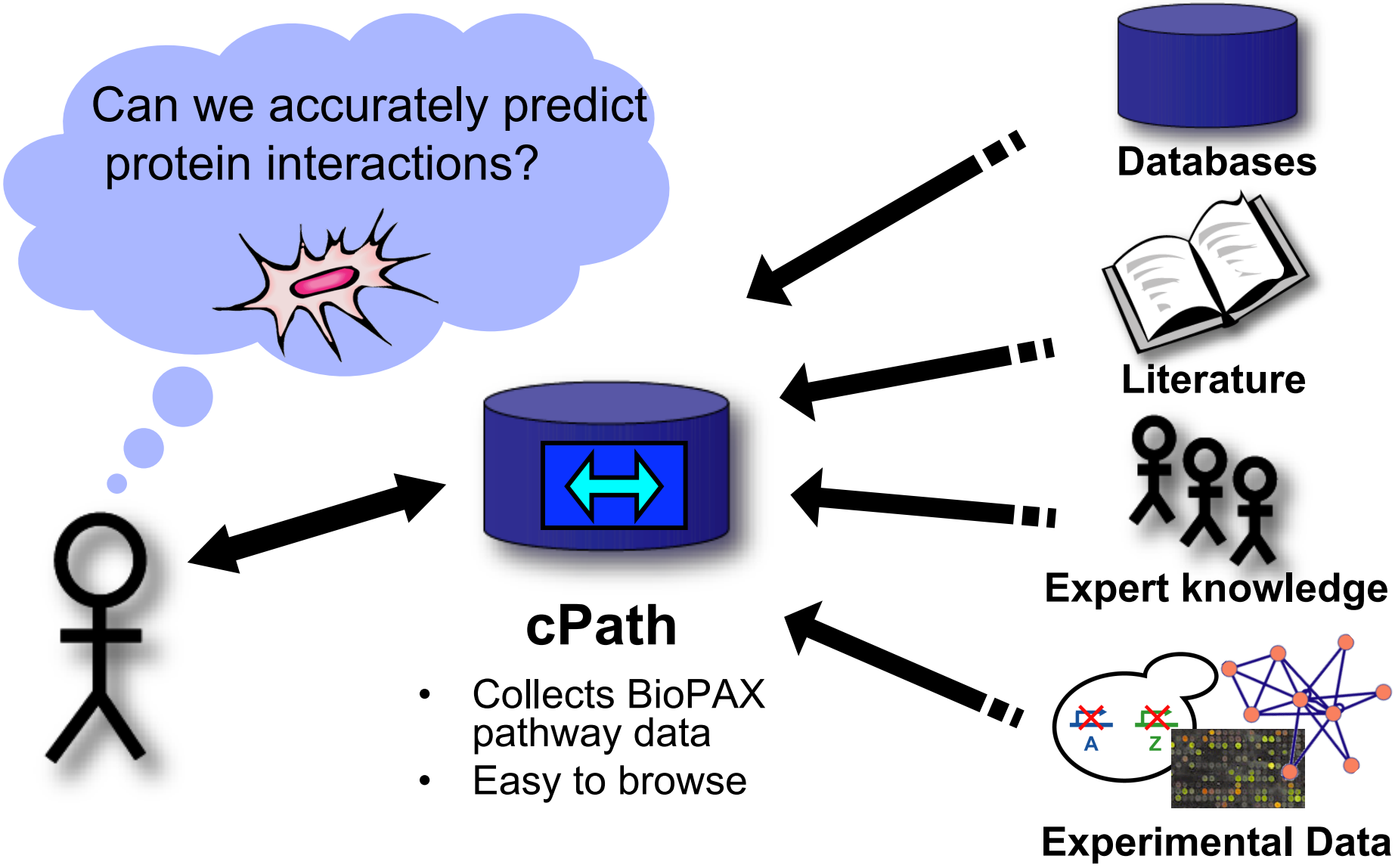
Grants/Support

- Department of Energy (Workshop)
- caBIG

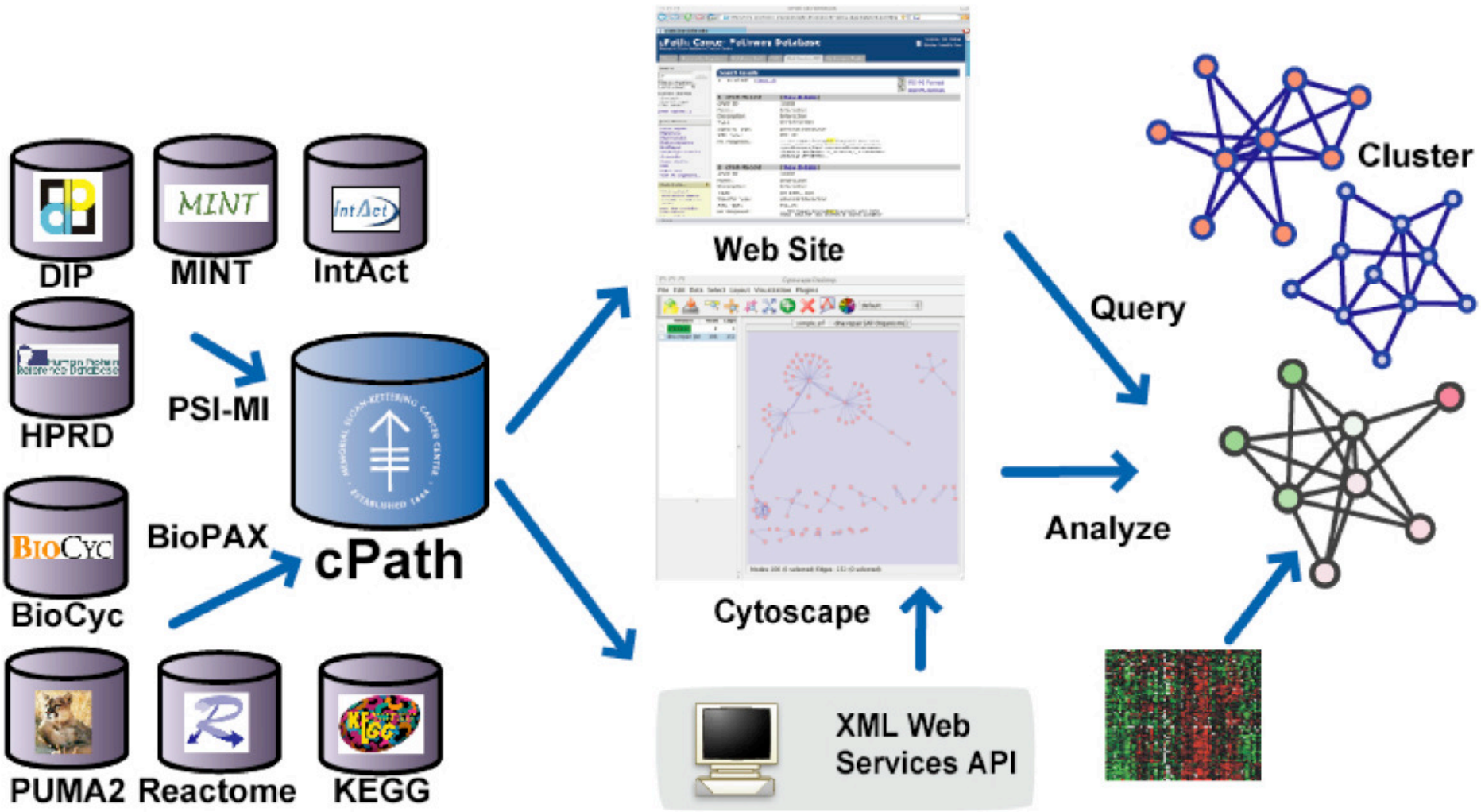
Using Pathway Information



Using Pathway Information



cPath Pathway Database Software



cPath Key Features

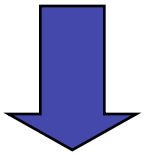
- Identifier mapping system e.g. proteins
- Scalable pathway data aggregation
- Simple web interface for browse and query
- Standard web service API for application communication
- 100% open source
  - Java, Tomcat, MySQL, Lucene, Struts, YUI
- Local installation and customization

http://cbio.mskcc.org/cpath

Cerami EG, Bader GD, Gross BE, Sander C.
BMC Bioinformatics. 2006 Nov 13;7:497

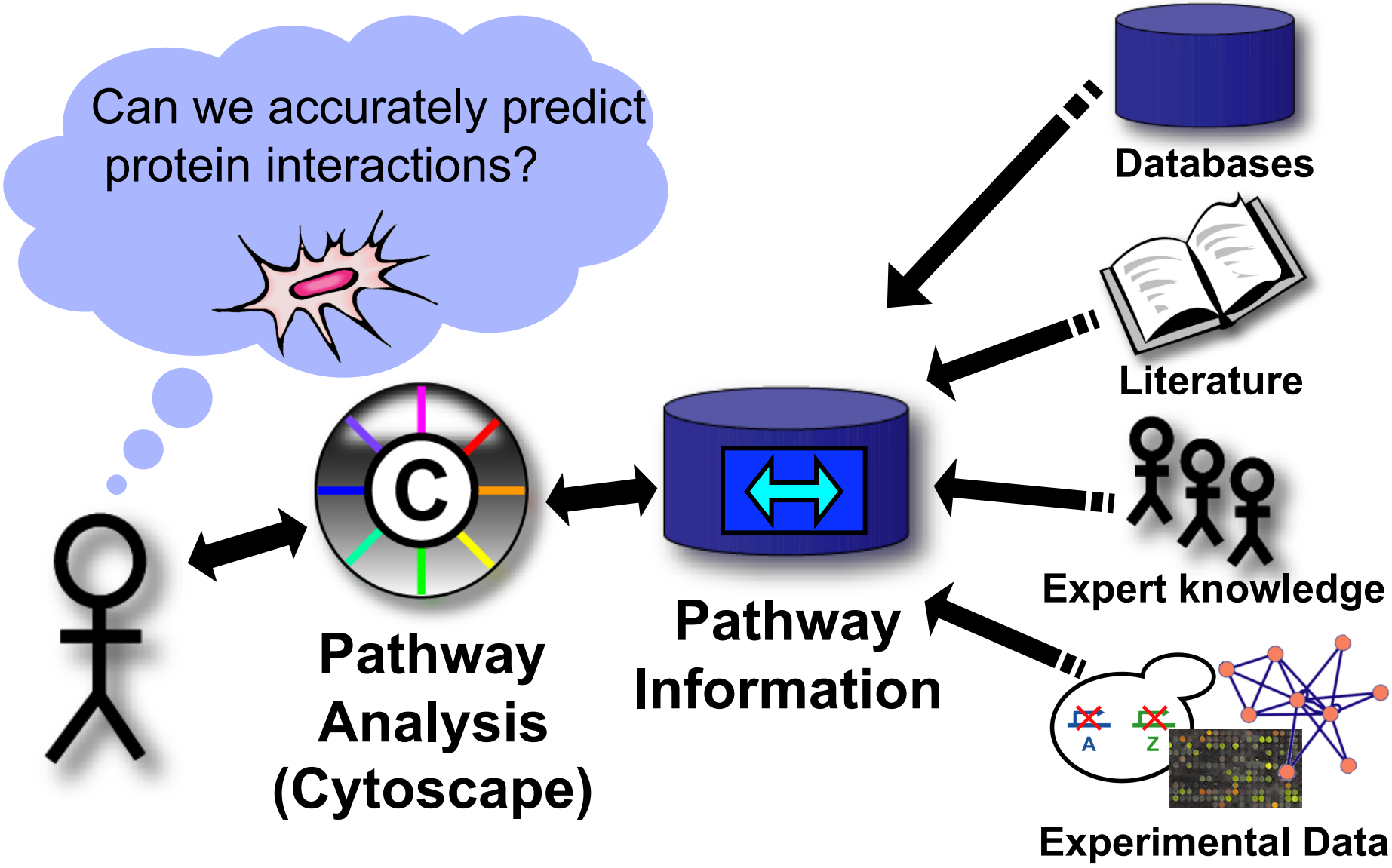
cPath web service API

- Queried by URL (RESTful architecture)
- getPathway, getNeighbors, getPathwayList, search
- webservice.do?cmd=get_pathway_list&version=2.0&q=O14763&input_id_type=UNIPROT



Database:ID	Pathway_Name	Pathway_Database_Name	Internal_ID
UNIPROT:O14763	Apoptosis	REACTOME	579
UNIPROT:O14763	Extrinsic Pathway for Apoptosis	REACTOME	580
UNIPROT:O14763	Death Receptor Signalling	REACTOME	581
UNIPROT:O14763	FasL/ CD95L signaling	REACTOME	582
UNIPROT:O14763	TRAIL signaling	REACTOME	584
UNIPROT:O14763	Caspase-8 is formed from procaspase-8	REACTOME	585
UNIPROT:O14763	Activation of Pro-Caspase 8	REACTOME	586

Using Pathway Information

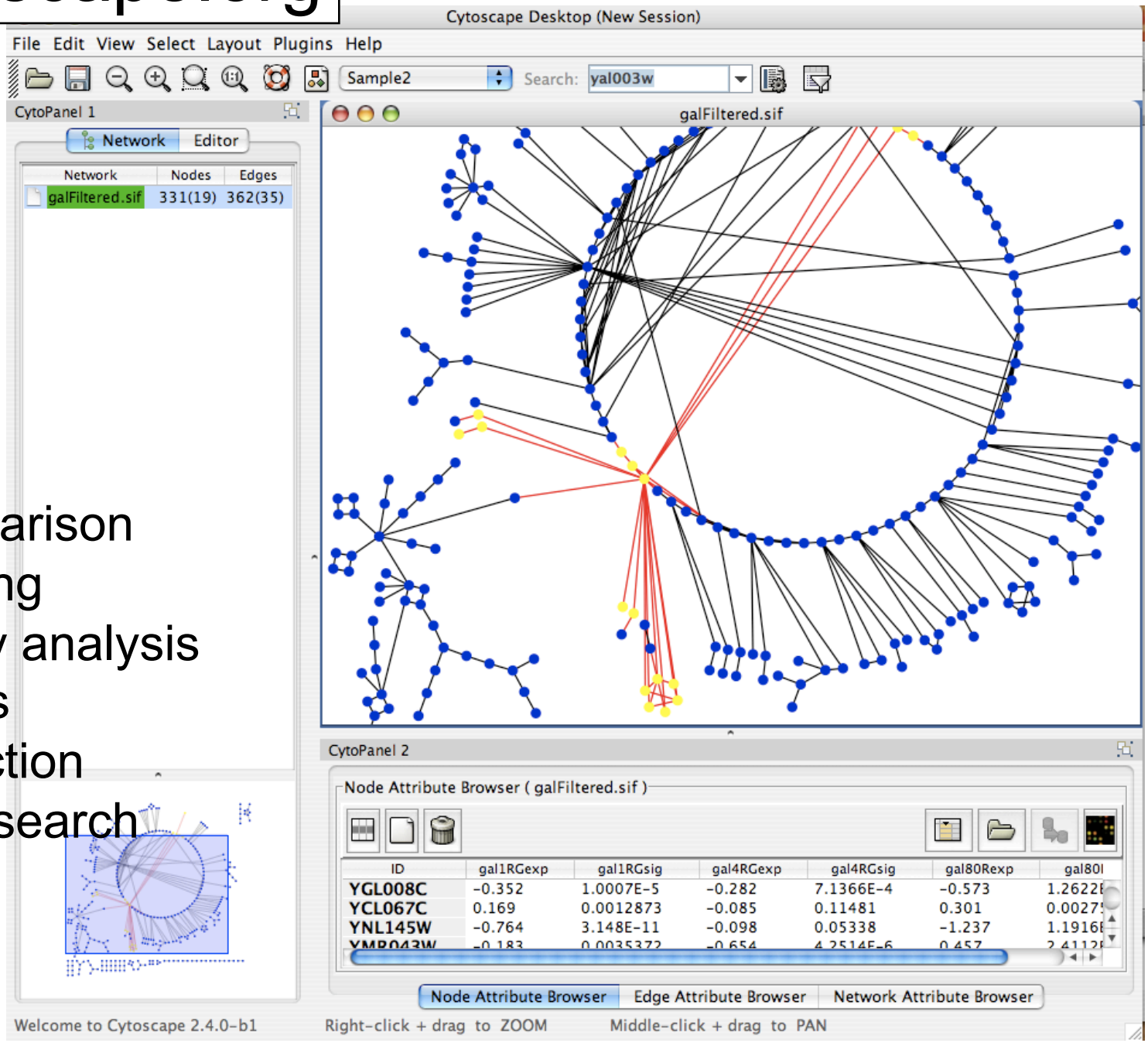


http://cytoscape.org

Network visualization and analysis

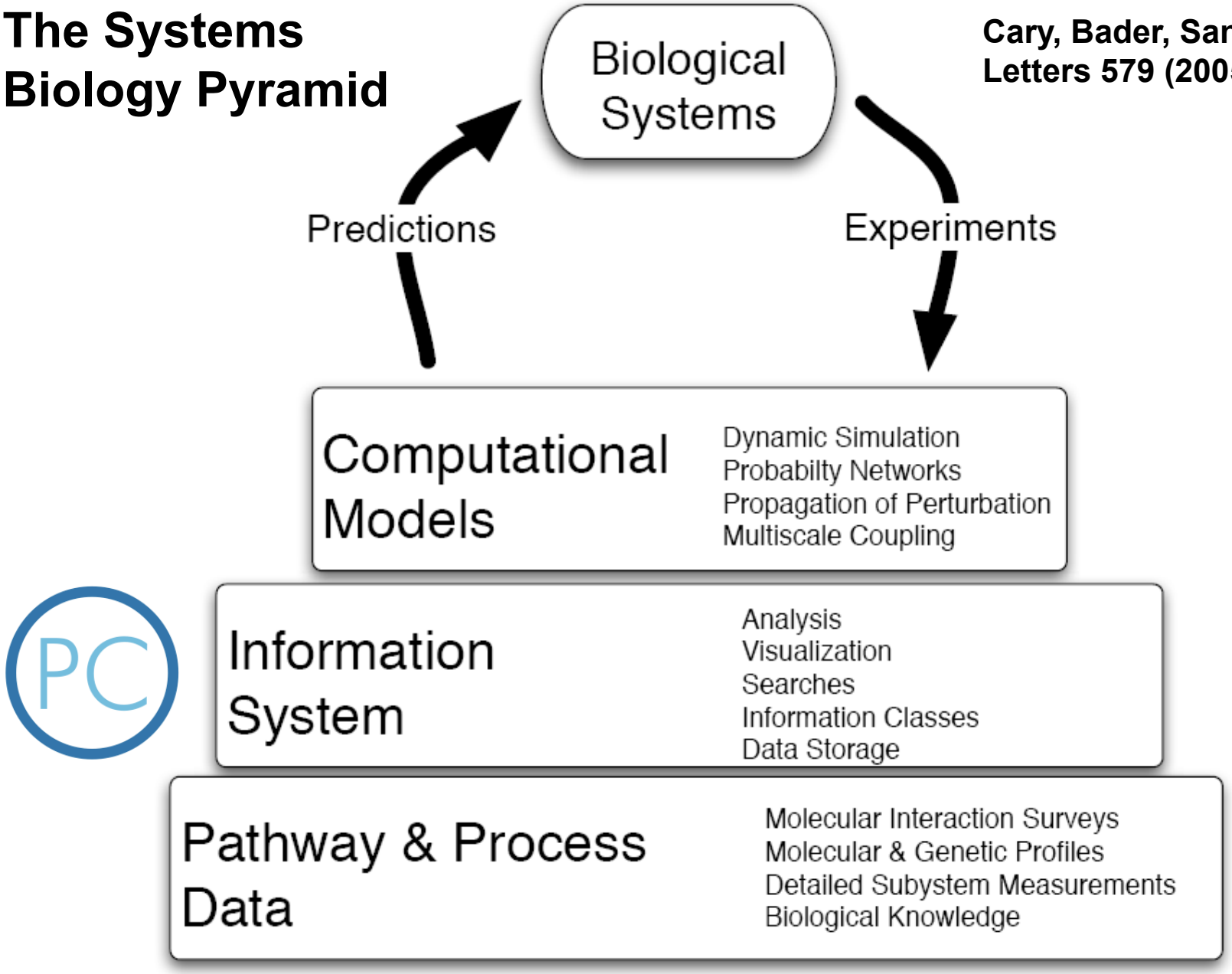
Pathway comparison
Literature mining
Gene Ontology analysis
Active modules
Complex detection
Network motif search

UCSD, ISB,
Agilent,
MSKCC,
Pasteur

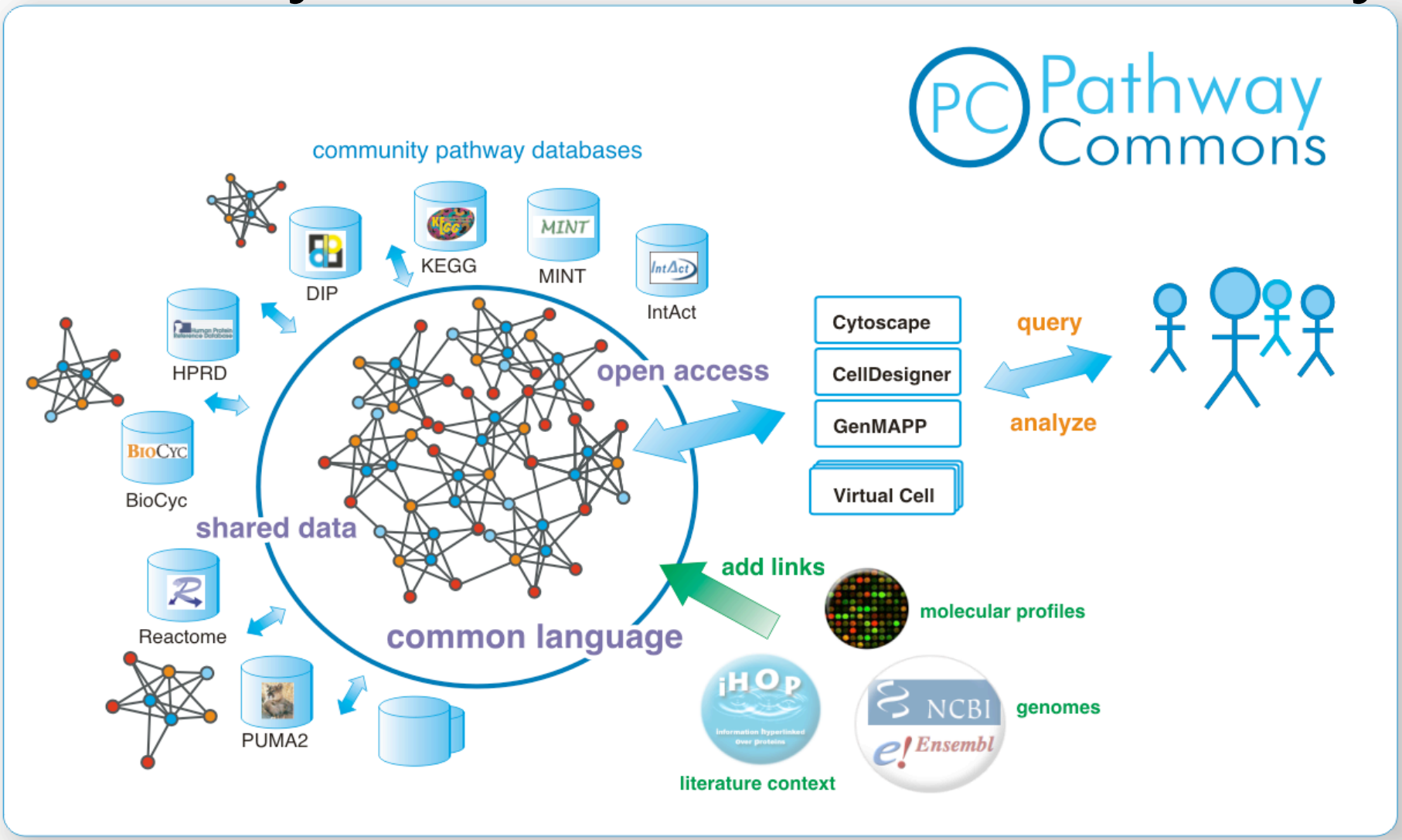


The Systems Biology Pyramid

Cary, Bader, Sander, FEBS Letters 579 (2005) 1815-20



Pathway Commons: A Public Library

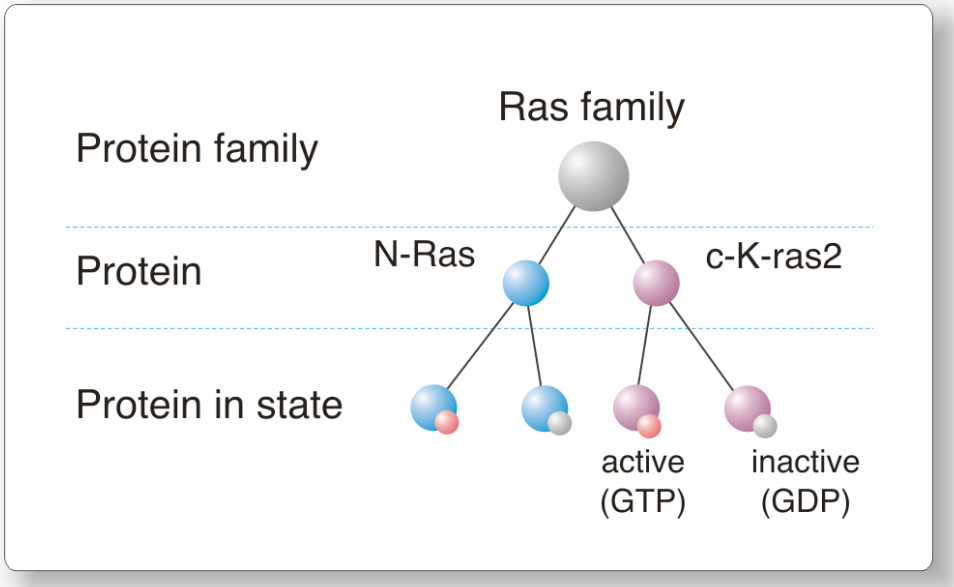


- Books: Pathways
- Lingua Franca: BioPAX OWL
- Index: cPath pathway database software
- Translators: translators to BioPAX

- Open access, free software
- No competition: Author attribution
- Aggregate ~ 20 databases in BioPAX format

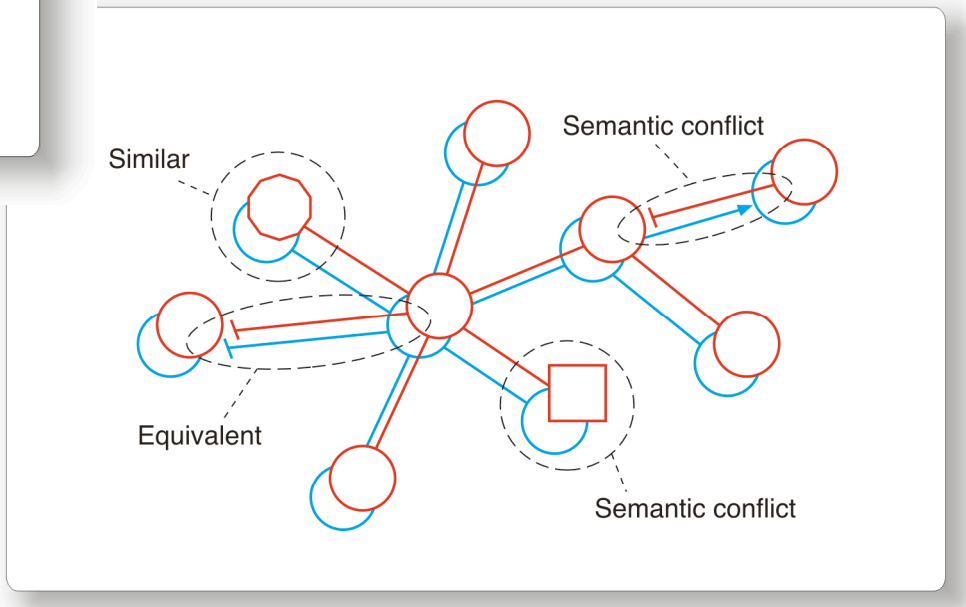
Towards an Integrated Cell Map

- Semantic pathway integration is very hard



Physical entities

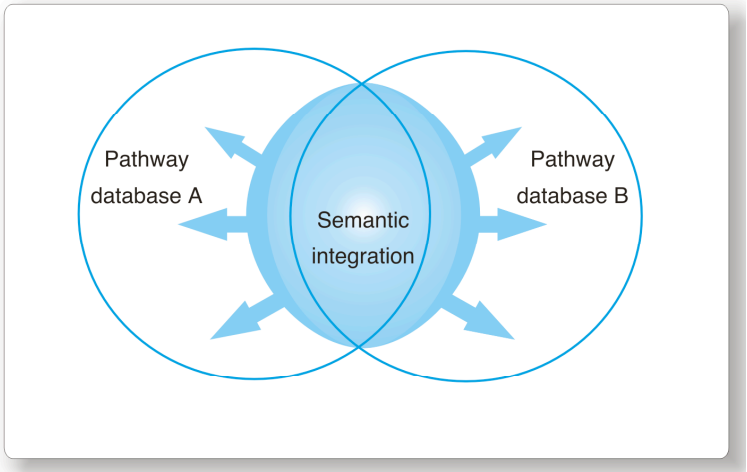
Relationships



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



http://pathwaycommons.org

Pathway Commons

Pathway Commons is a project for the scientific community. We welcome your feedback at: pc-info@pathwaycommons.org.

Home | Filter | Stats | FAQ | Cytoscape | About | Credits

Pathway Commons is a convenient point of access to biological pathway information collected from public pathway databases, which you can browse or search. Pathways include biochemical reactions, complex assembly, transport and catalysis events, and physical interactions involving proteins, DNA, RNA, small molecules and complexes. more...

Search Pathway Commons:

Search

To get started, enter a gene name, gene identifier or pathway name in the text box above. For example: p53, P38398 or mTOR.

To restrict your search to specific data sources or specific organisms, update your global filter settings.

Biologists: Browse and search pathways across multiple valuable public pathway databases.

Computational biologists: Download an integrated set of pathways in BioPAX format for global analysis.

Software developers: Build software on top of Pathway Commons using our soon-to-be released web service API. Download and install the cPath software to create a local mirror.

Pathway Commons currently contains the following data sources:

Cancer Cell Map, Release: 1.0 [19-May-06]
Browse

HumanCyc, Release: 10.5 [18-Sep-06]
Browse

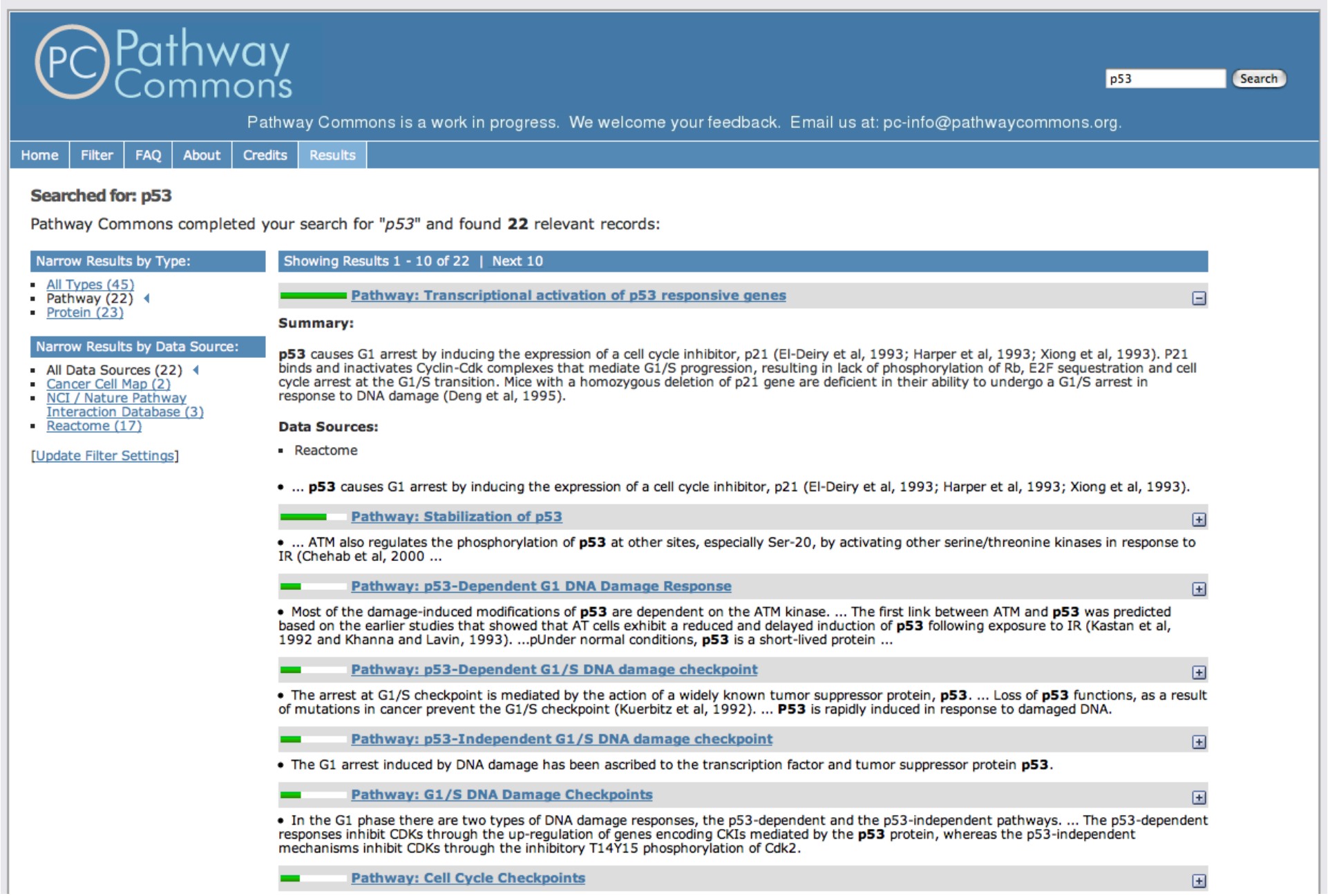
NCI / Nature Pathway Interaction Database [01-Jan-07]
Browse

Reactome, Release: 19 [16-Nov-06]
Browse

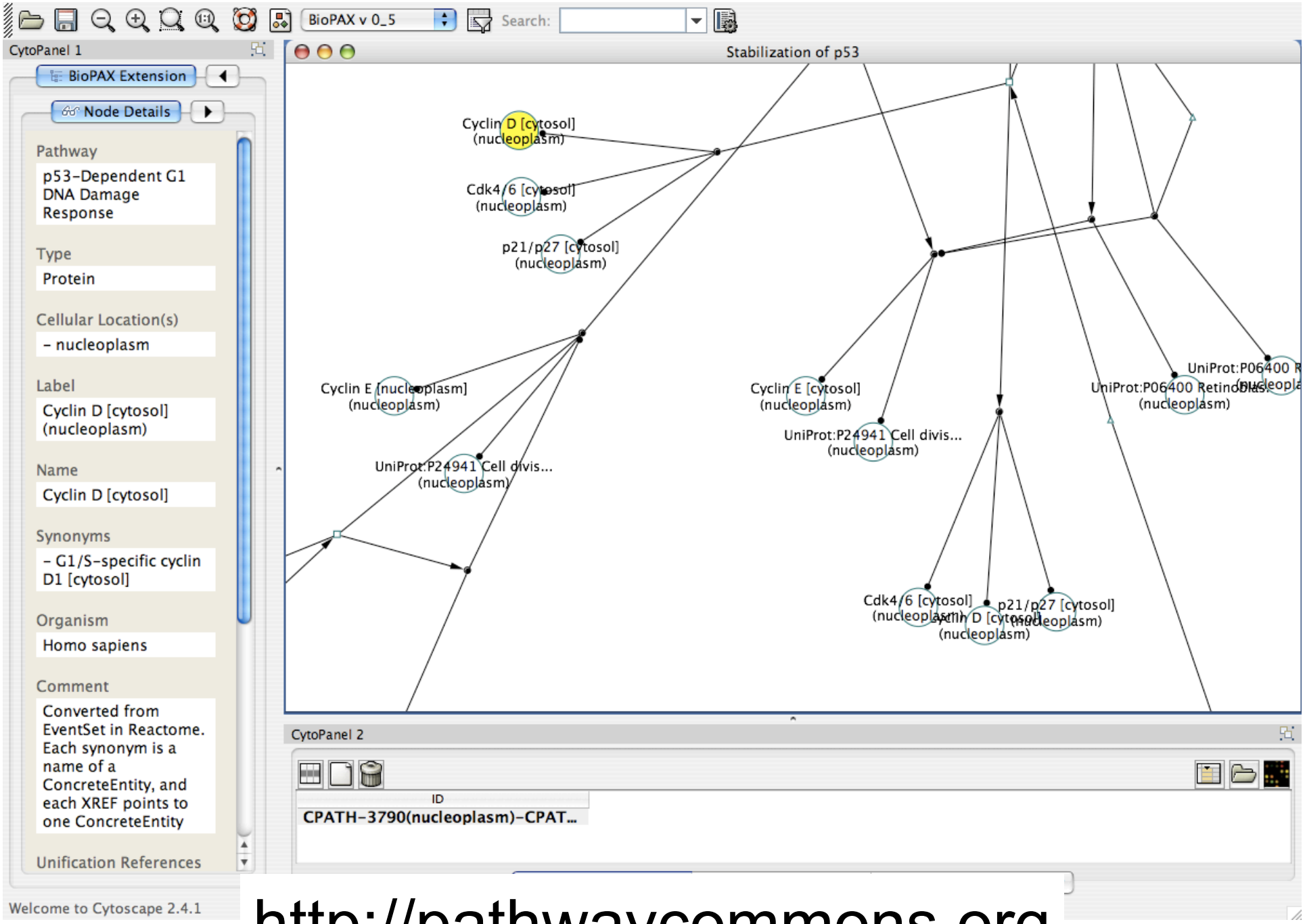
Pathway Commons Quick Stats:

Number of Pathways:	921
Number of Interactions:	9,924
Number of Physical Entities:	15,515
Number of Organisms:	10

Integration of additional data sources is planned in the near future. For a comprehensive directory of interaction and pathway databases, please refer to Pathguide.

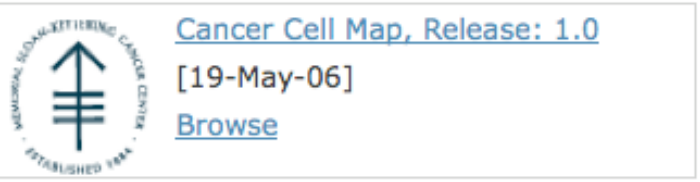


http://pathwaycommons.org



http://pathwaycommons.org

Pathway Commons Status



- More Databases
  - iHOP annotation
- Web service API under development
- Neighborhood visualization
- Cytoscape integration

Pathway Commons Quick Stats:

Number of Pathways:	1,111
Number of Interactions:	258,282
Number of Physical Entities:	50,183
Number of Organisms:	445

Open Challenges

- Data: Author entry systems
  - From individual publications (The Cashew Prize)
  - For pathways (review)
  - Curator tools (advanced)
- Semantic integration (ID resolution)
- Visualization
  - Pathway diagrams (SBGN)
  - Automated layout
- Algorithms for compound graphs
- Linking discrete and dynamic representations
  - Including use by modelers

Acknowledgements

Pathway Commons

Chris Sander
Ethan Cerami
Ben Gross
Emek Demir
Robert Hoffmann
(Robert Sheridan)

Cytoscape

Trey Ideker (UCSD)
Ryan Kelley, Kei Ono, Mike Smoot, Peng Liang Wang
(Nerius Landys, Chris Workman, Mark Anderson, Nada Amin, Owen Ozier, Jonathan Wang)

Lee Hood (ISB)
Sarah Killcoyne
(Iliana Avila-Campillo, Rowan Christmas, Andrew Markiel, Larissa Kamenkovich, Paul Shannon)

Benno Schwikowski (Pasteur)
Melissa Cline, Tero Aittokallio

Chris Sander (MSKCC)
Ethan Cerami, Ben Gross (Robert Sheridan)

Bader Lab
G2N
Chris Tan
David Gfeller
Lianet Lopez
Moyez Dharsee
Shirley Hui

MP
Anastasija Baryshnikova
Iain Wallace
Laetitia Morrison

ACM
Daniele Merico
Ruth Isserlin
Vuk Pavlovic

Annette Adler (Agilent)
Allan Kuchinsky, Mike Creech, (Aditya Vailaya),

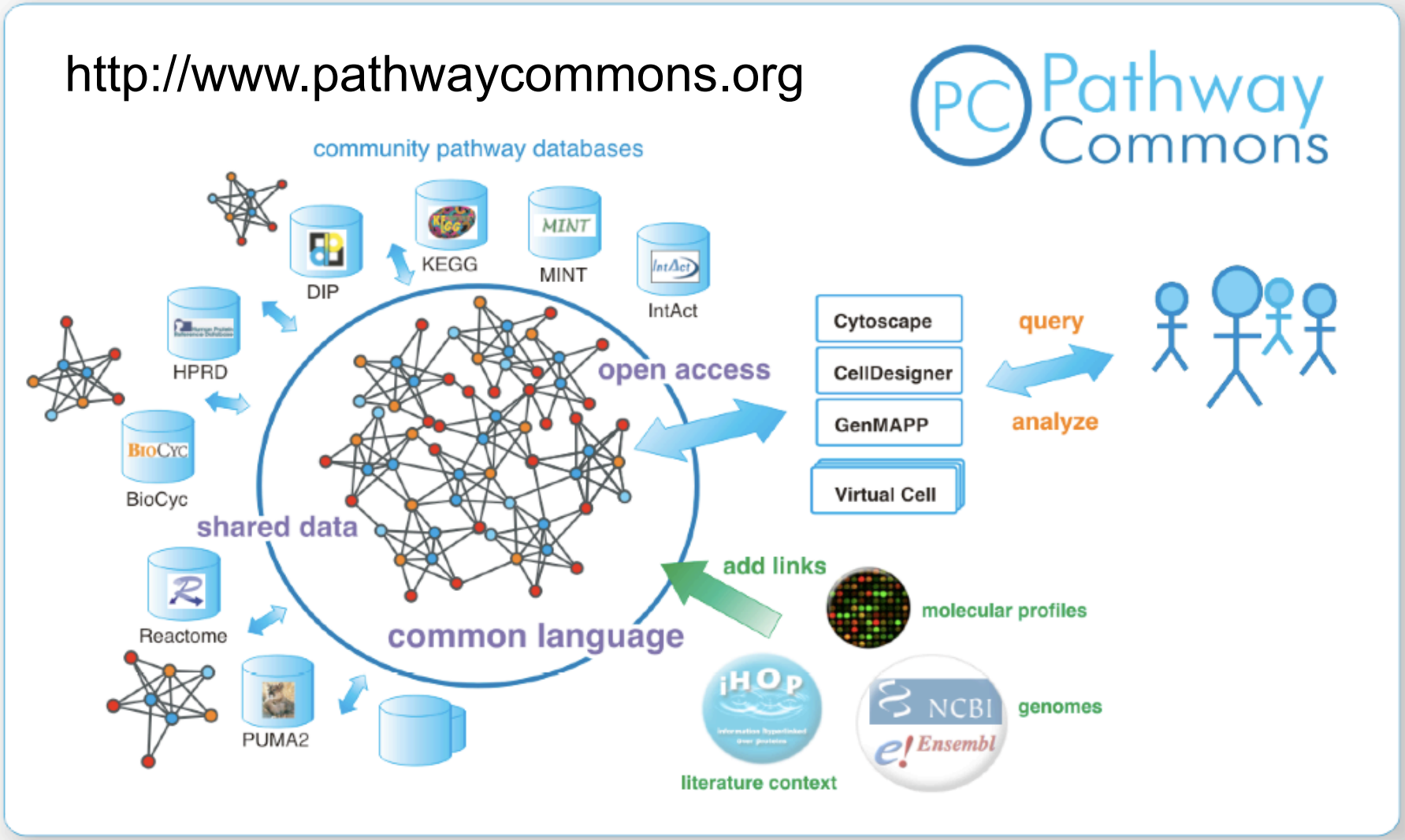
Bruce Conklin (UCSF)
Alex Pico, Kristina Hanspers
John ‘Scooter’ Morris (Ferrin lab, UCSF)

Funding

NIH, Genome Canada
Canada Foundation for Innovation/ORF

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