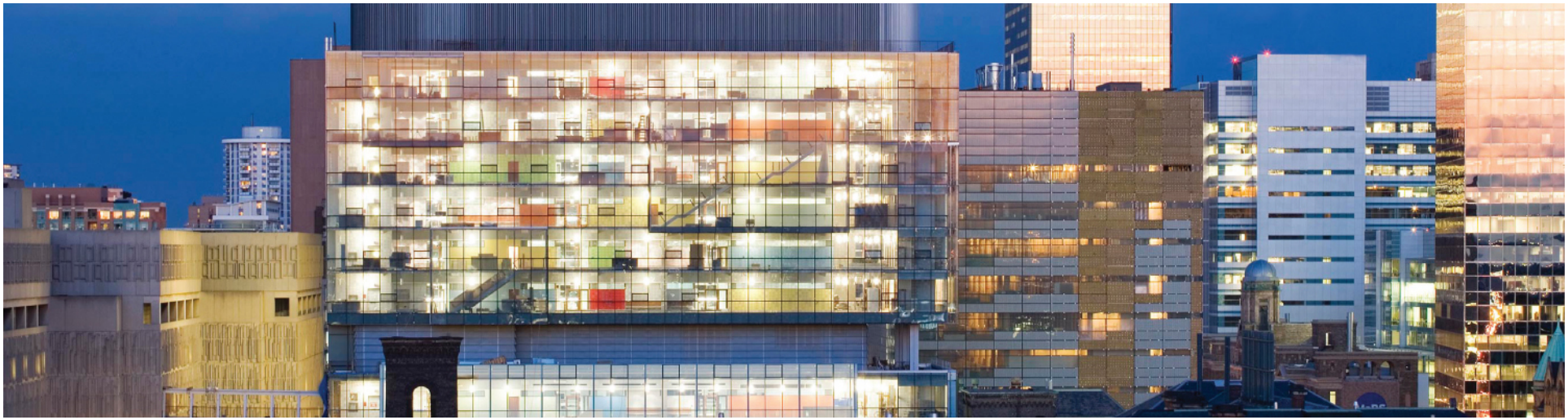
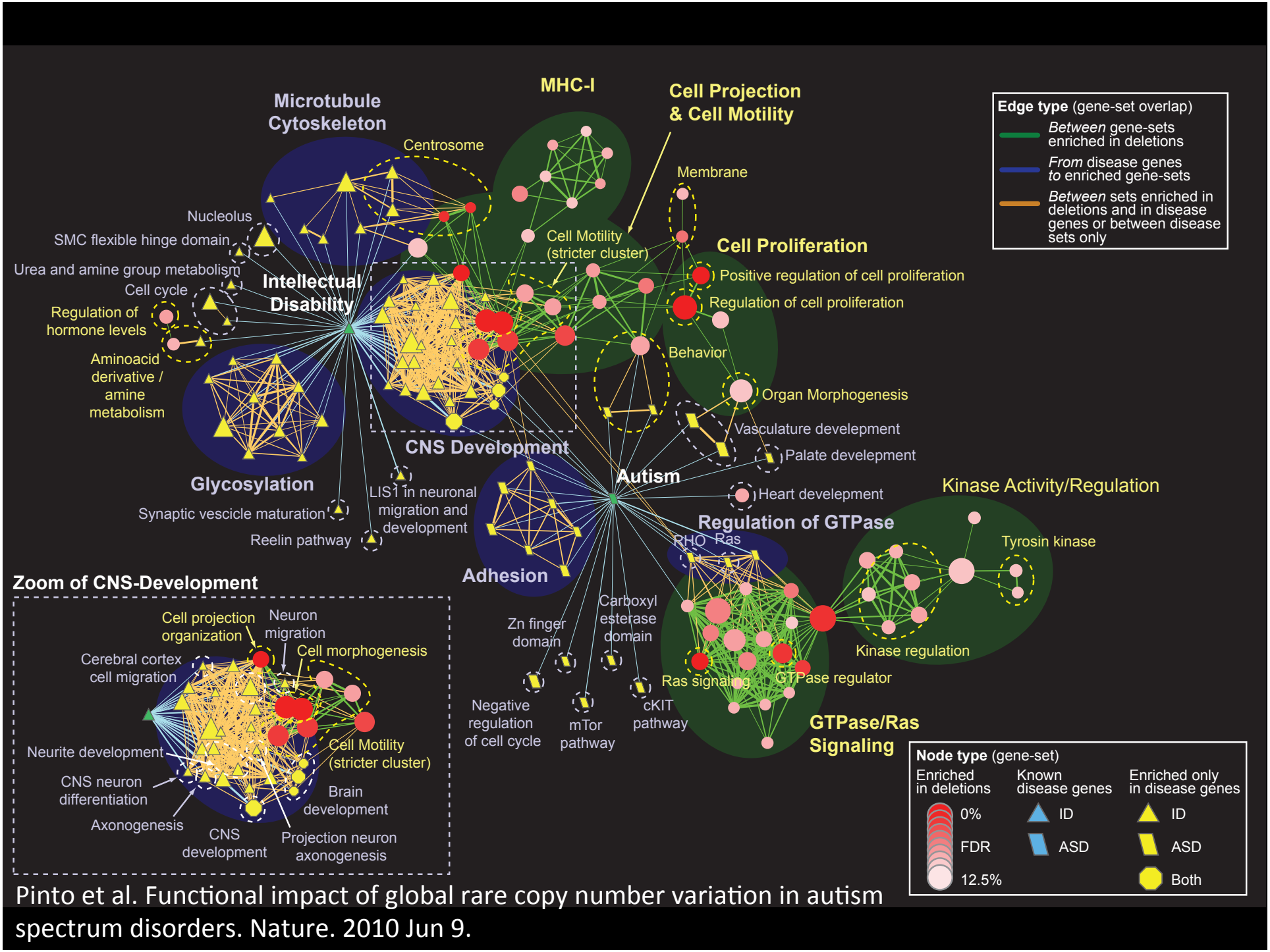
Pathway analysis for genomics

Gary Bader
Oct.23.2012 – FGED Toronto

Donnelly Centre
for Cellular + Biomolecular Research

UNIVERSITY OF TORONTO





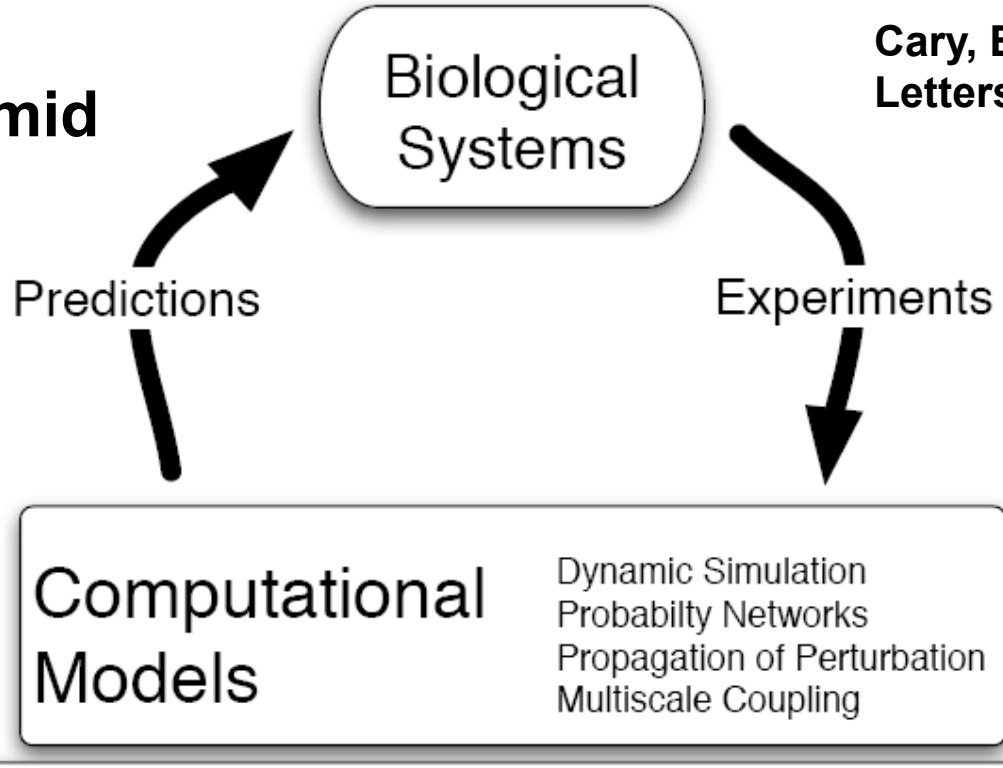
Pinto et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010 Jun 9.

Correlation to Causation

- GWAS: find genetic markers correlated with disease – powerful approach, but:
  - genomics reduces statistical power (>multiple testing correction with >SNPs)
  - rare variants = more samples
- Associate pathways to increase power
  - Fewer pathways, organize many rare variants (damaging the system causes the disease)
- Use pathway knowledge to identify potential disease causes

The Systems Biology Pyramid

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

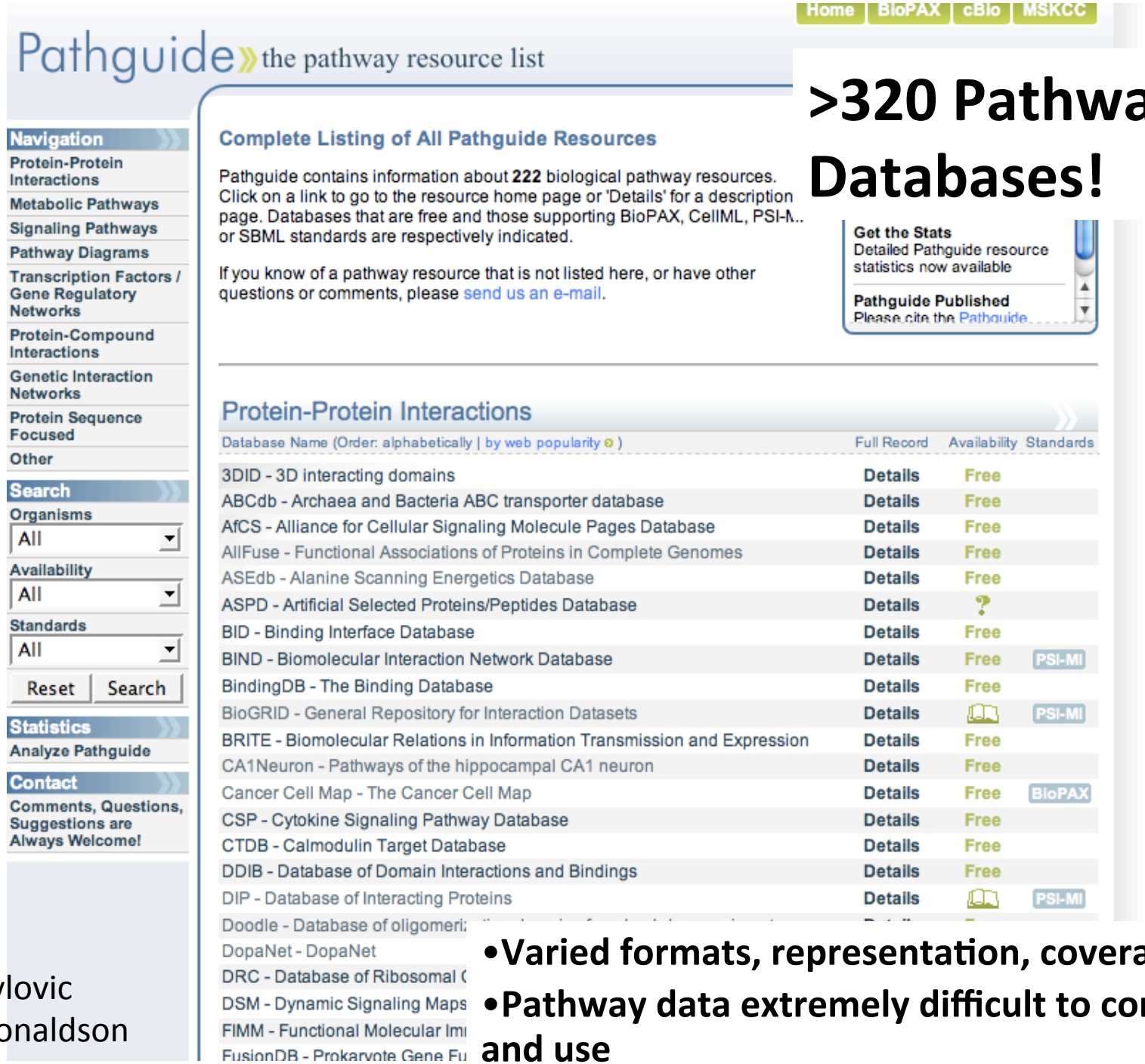


Cytoscape
cPath2

Pathway Commons
BioPAX

Chris Sander, MSKCC

http://pathguide.org



>320 Pathway Databases!

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

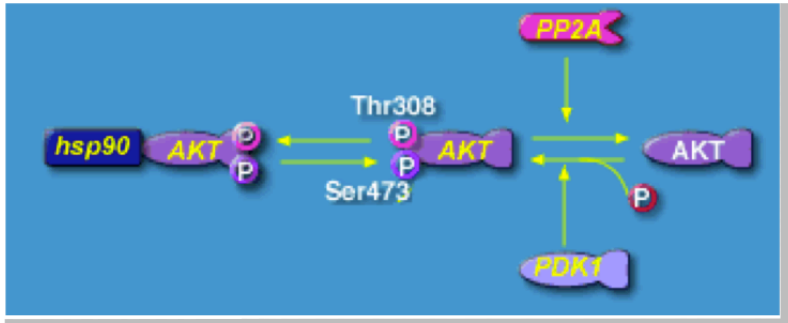
Vuk Pavlovic
Sylva Donaldson

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

www.biopax.org

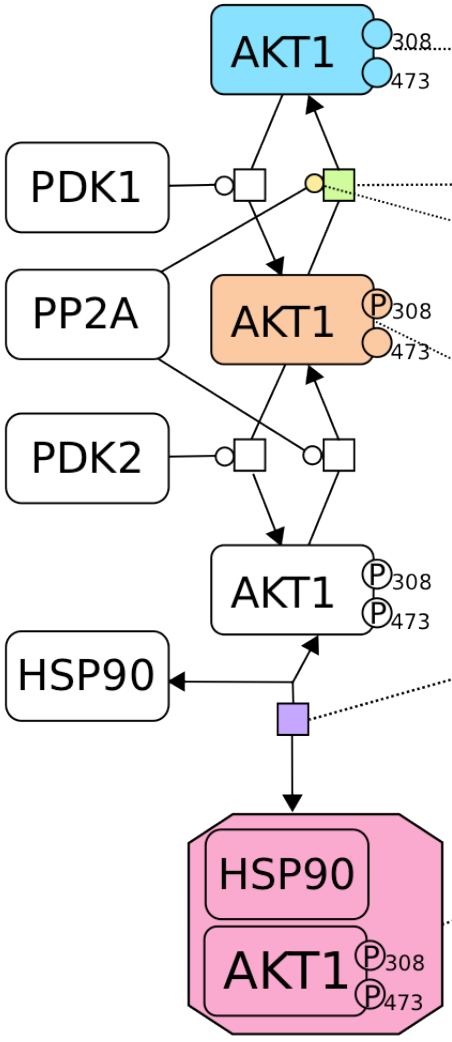
BioCarta



rAKT1 is a *ProteinReference*
has *standard-name* "AKT1"
has *name* "PKB"
has *xref* **Uniprot-P31749**

p@308 is a *ModificationFeature*
has *featureLocation* **AKT1-308**
has *modificationType*
phosphorylation

SBGN.org



BioPAX

AKT1.1 is a *Protein*
has *proteinReference* **rAKT1**
has *notFeature* **p@308**
has *notFeature* **p@473**

reaction1 is a *BiochemicalReaction*
has *left* **AKT1.2**
has *right* **AKT1.1**
is *left-to-right.*

catalysis1 is a *Catalysis*
has *controller* **PP2A.1**
has *controlled* **reaction1**
has *direction irr-left-to-right*

AKT1.2 is a *Protein*
has *proteinReference* **rAKT1**
has *feature* **p@308**
has *notFeature* **p@473**

assembly1 is a *ComplexAssembly*
has *left* **HSP90.1**
has *left* **AKT1.3**
has *right* **complex1**
is *reversible*

complex1 is a *Complex*
has *component* **AKT1.4**
has *component* **HSP90.2**

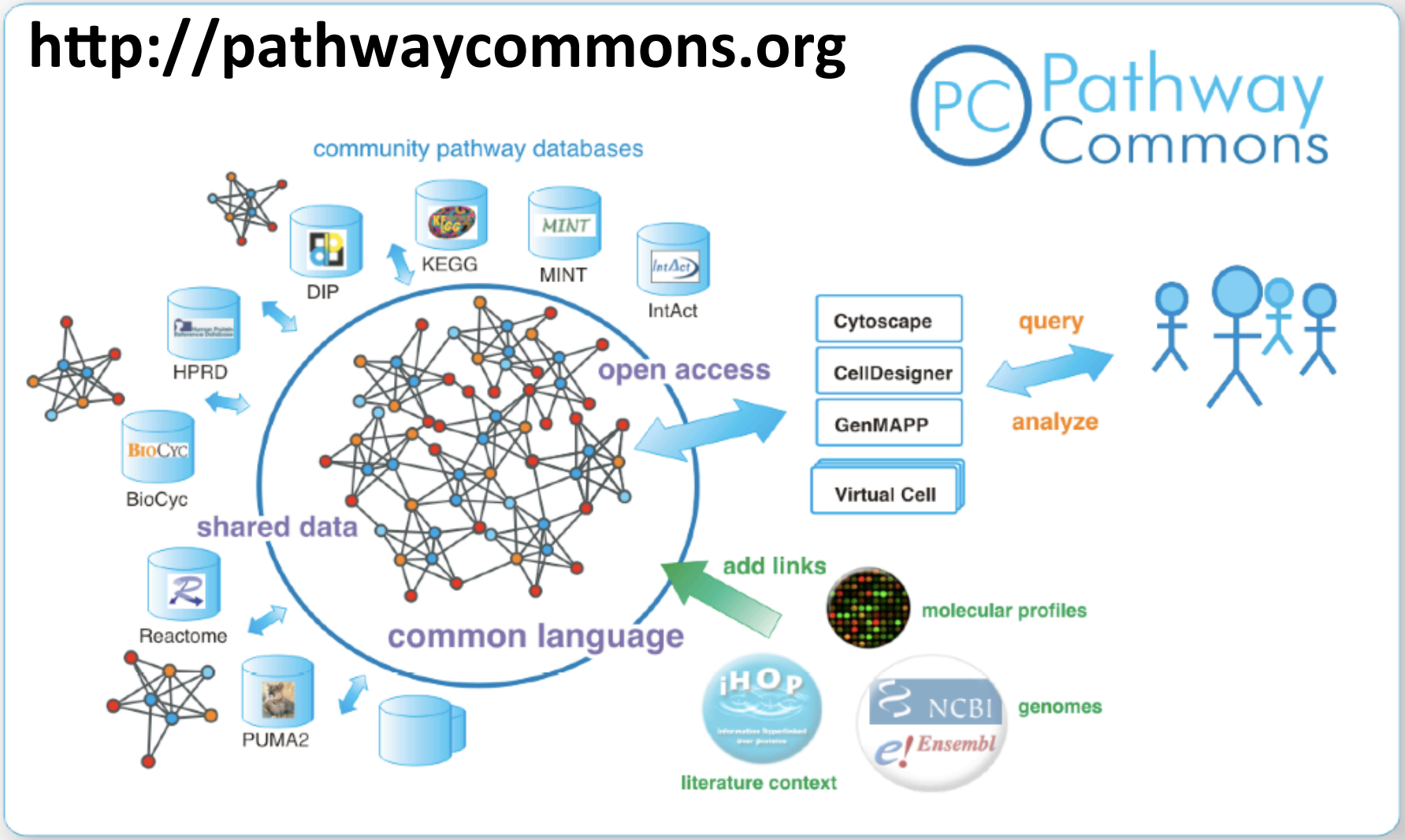
HSP90.2 is a *Protein*
has *proteinReference* **rHSP90**
is *boundTo* **AKT1.4**

AKT1.4 is a *Protein*
has *proteinReference* **rAKT1**
has *feature* **p@308**
has *feature* **p@473**
is *boundTo* **HSP90.2**

Emek Demir

Aim: Convenient Access to Pathway Information

http://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

Pathway Commons: cPath^2

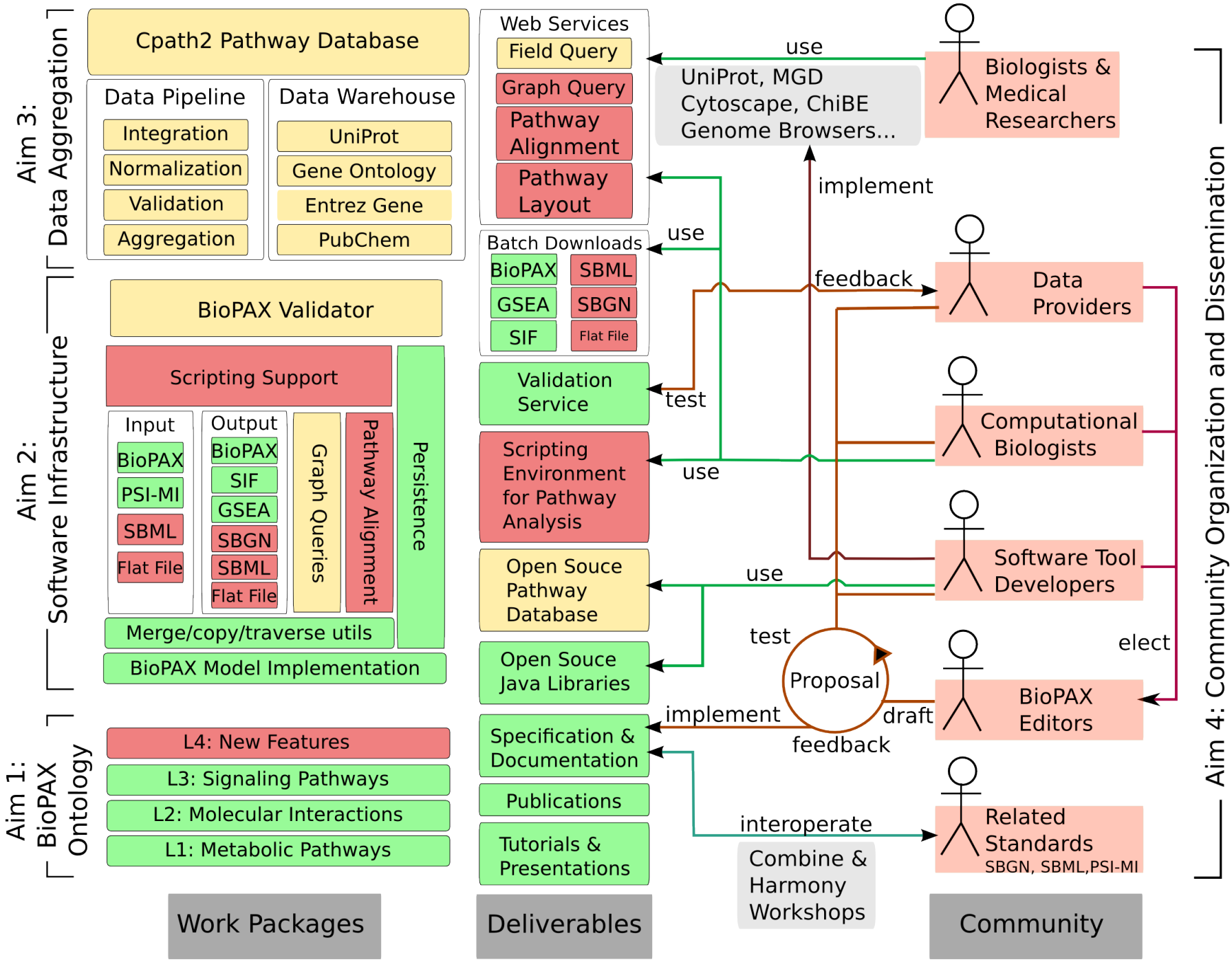
- http://www.pathwaycommons.org/pc2/

Web Service API:

You can programmatically access the data within Pathway Commons using the Pathway Commons Web Service Application Programming Interface (API). This page provides a reference guide to help you get started.

1. Command: SEARCH
2. Command: GET
3. Command: GRAPH
4. Command: TRAVERSE
5. Command: TOP_PATHWAYS
6. Command: HELP

Emek Demir, Igor Rodchenkov, Chris Sander Ozgun Babur, Arman Aksoy, Onur Sumer, Ethan Cerami, Ben Gross

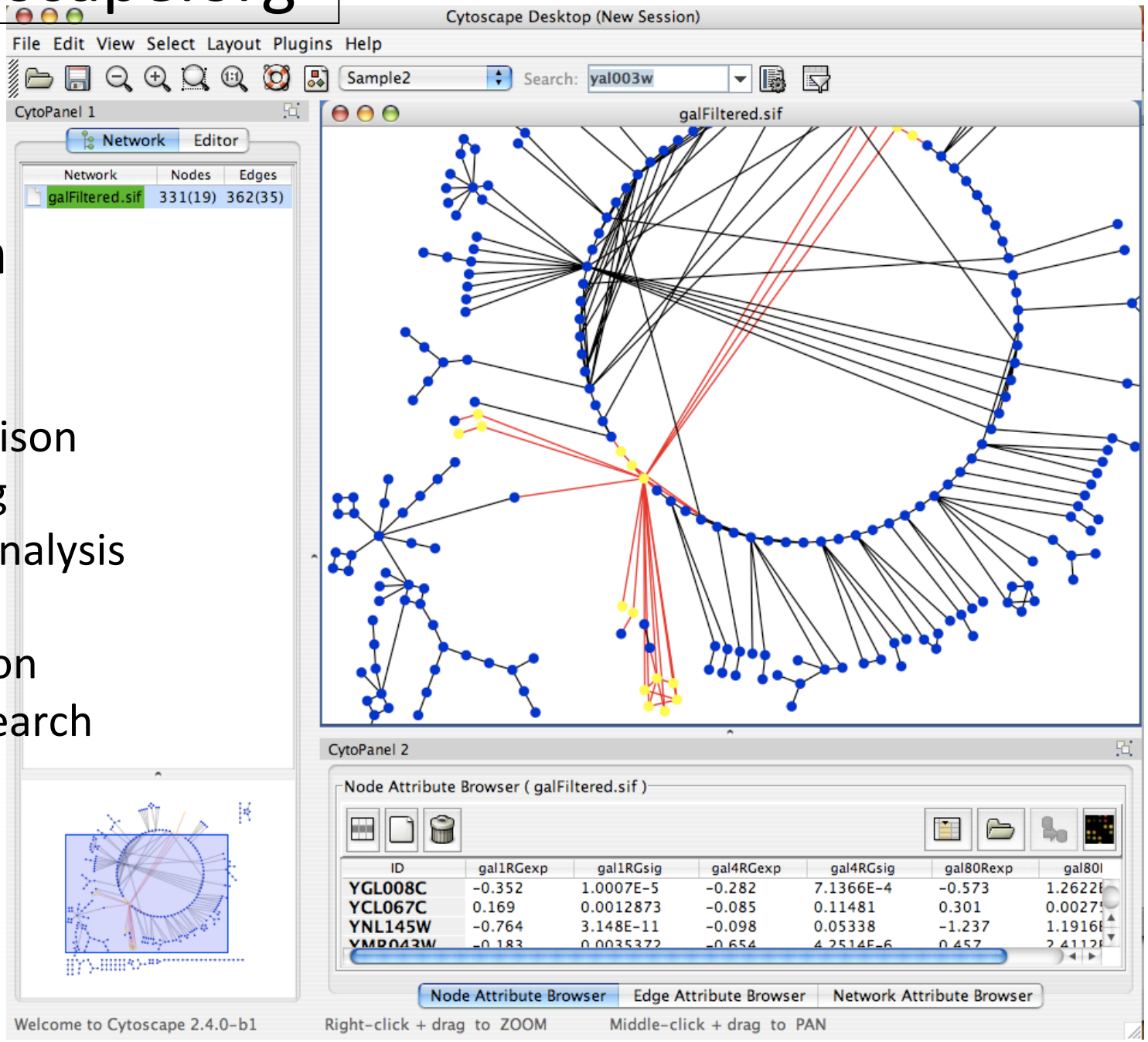

Aim 3: Data Aggregation
Aim 2: Software Infrastructure
Aim 1: BioPAX Ontology
Aim 4: Community Organization and Dissemination
Cpath2 Pathway Database
Data Pipeline
Integration
Normalization
Validation
Aggregation
Data Warehouse
UniProt
Gene Ontology
Entrez Gene
PubChem
BioPAX Validator
Scripting Support
Input
BioPAX
PSI-MI
SBML
Flat File
Output
BioPAX
SIF
GSEA
SBGN
SBML
Flat File
Graph Queries
Pathway Alignment
Persistence
Merge/copy/traverse utils
BioPAX Model Implementation
L4: New Features
L3: Signaling Pathways
L2: Molecular Interactions
L1: Metabolic Pathways
Work Packages
Web Services
Field Query
Graph Query
Pathway Alignment
Pathway Layout
Batch Downloads
BioPAX
SBML
GSEA
SBGN
SIF
Flat File
Validation Service
Scripting Environment for Pathway Analysis
Open Souce Pathway Database
Open Souce Java Libraries
Specification & Documentation
Publications
Tutorials & Presentations
Deliverables
use
UniProt, MGD
Cytoscape, ChiBE
Genome Browsers...
Biologists & Medical Researchers
implement
use
feedback
Data Providers
test
Computational Biologists
use
Software Tool Developers
use
test
elect
Proposal
draft
BioPAX Editors
implement
feedback
interoperate
Related Standards
SBGN, SBML,PSI-MI
Combine & Harmony Workshops
Community

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, UToronto, U Texas



Cytoscape 3

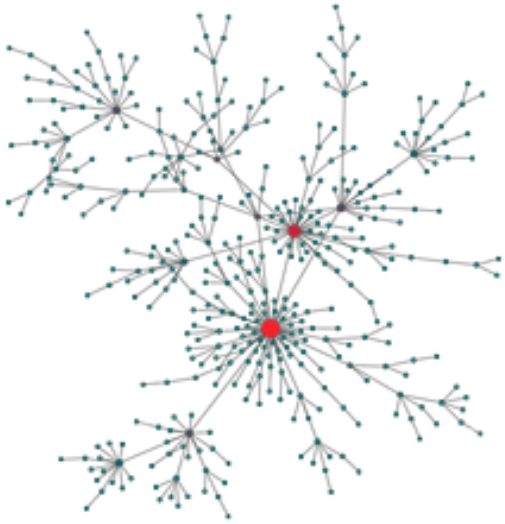
- Complete re-architecture: OSGi – everything is an app
- Enables future features:
  - More stable and powerful APIs
  - Scripting, macros, recordable history, better undo/redo
  - Command line mode, good for use on compute clusters
  - Interactive control from other scripting languages e.g. R
- Fixing bugs and porting plugins
- 3.0 beta now available
  - Mirror functionality in 2.8
  - Encourage plugin to app porting
  - http://www.cytoscape.org/cy3.html

National Resource for Network Biology

Mission Statement

The aim of the National Resource for Network Biology (NRNB) is to provide a freely available, open-source suite of software technology that broadly enables network-based visualization, analysis, and biomedical discovery for NIH-funded researchers. This software is enabling researchers to assemble large-scale biological data into models of networks and pathways and to use these networks to better understand how biological systems operate under normal conditions and how they fail in disease.

The National Resource for Network Biology is an NIH National Center for Research Resources (**NCRR**) Biomedical Technology Research Center (**BTRC**), organized around the following key components: Technology Research and Development, Driving Biomedical Projects, Outreach, Training and Dissemination of Tools. NRNB is funded by RR031228.



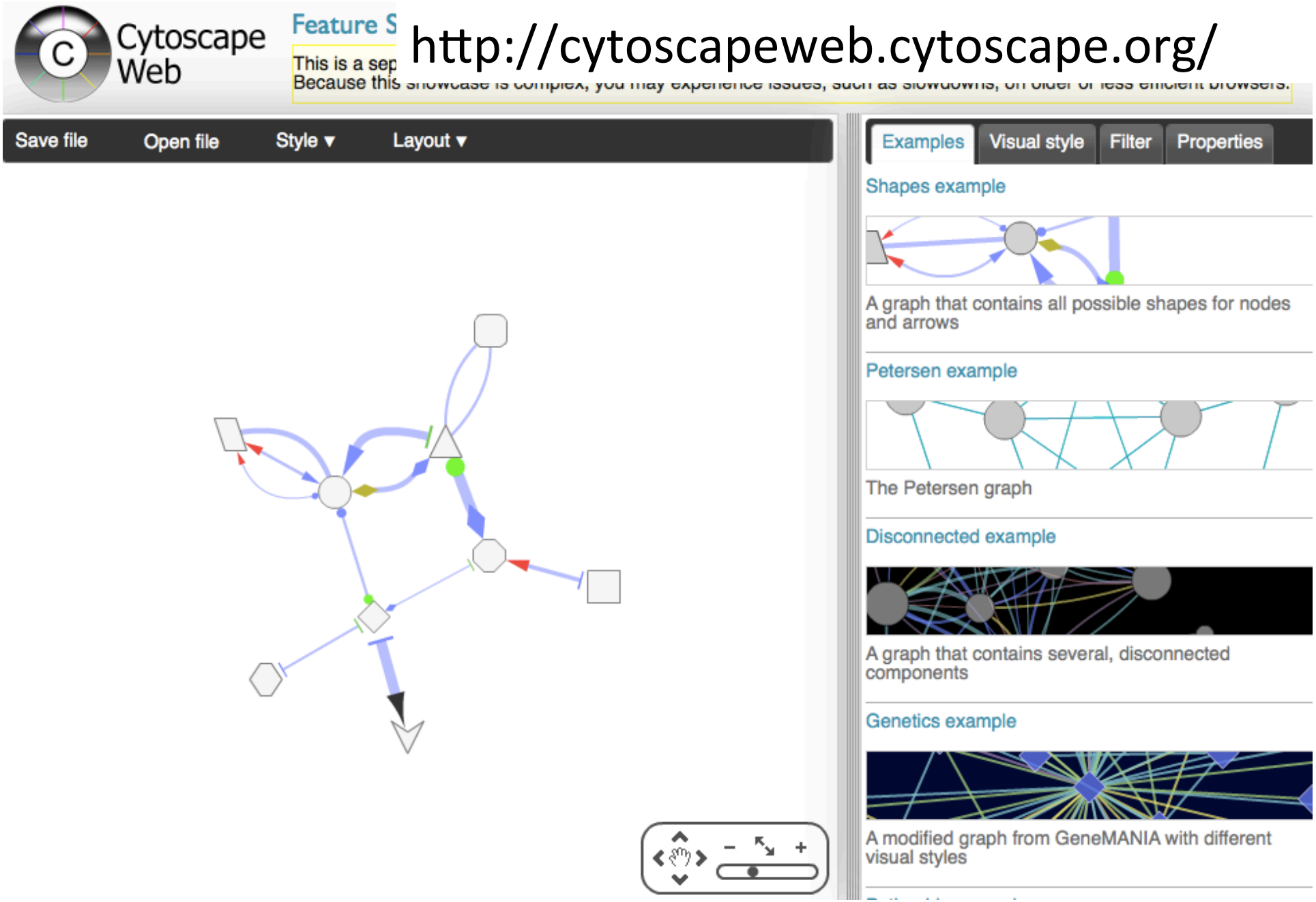
What is Network Biology?

Collaborate with NRNB

View ongoing collaborations

Annual Report

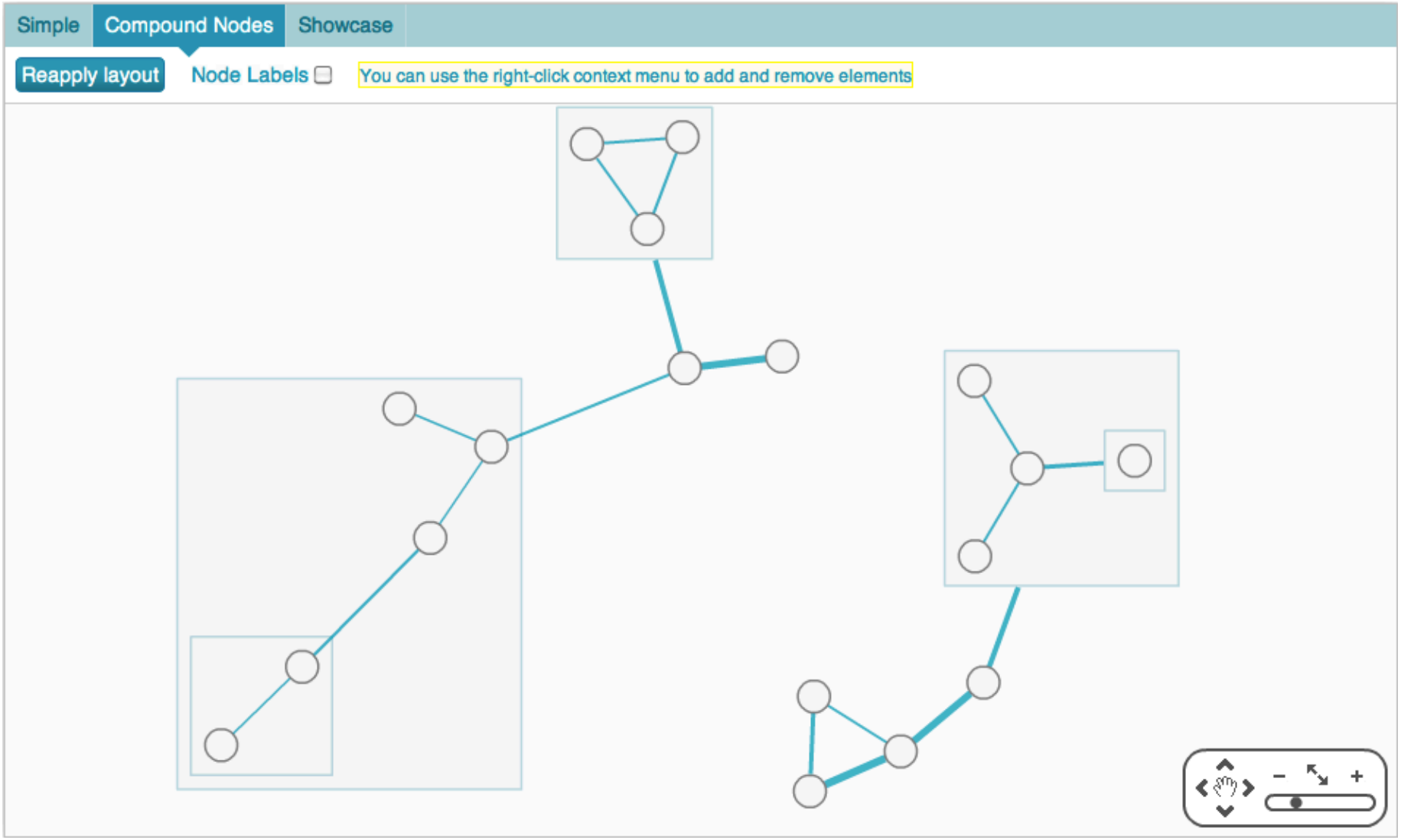
http://cytoscapeweb.cytoscape.org/



Christian Tannus-Lopes, Max Franz, Yue Dong

Onur Sumer
Ugur Dogrusoz
Bilkent, Ankara

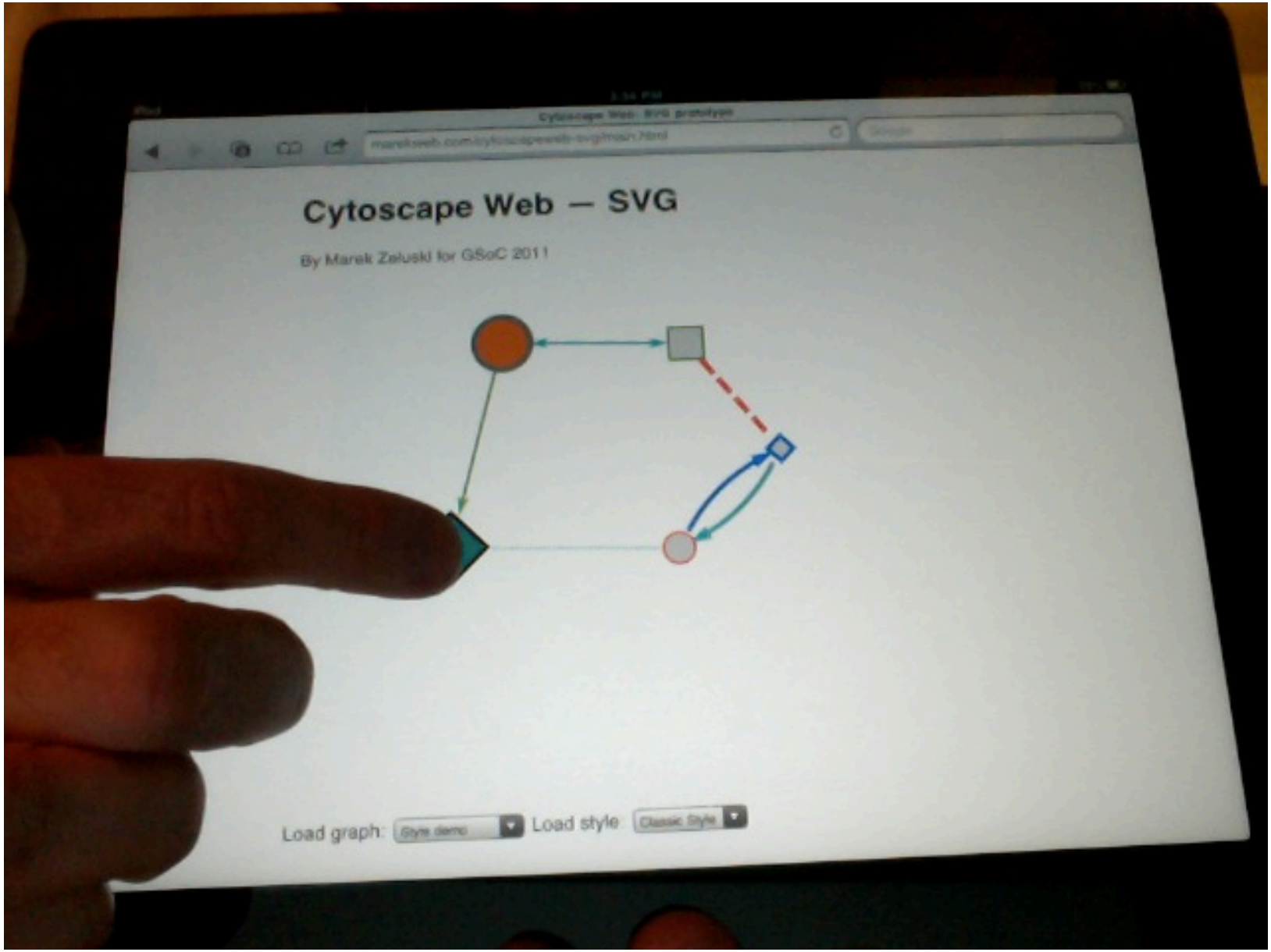
Compound Nodes



http://cytoscapeweb.cytoscape.org/demos/compound

cytoscape.github.com/cytoscape.js/

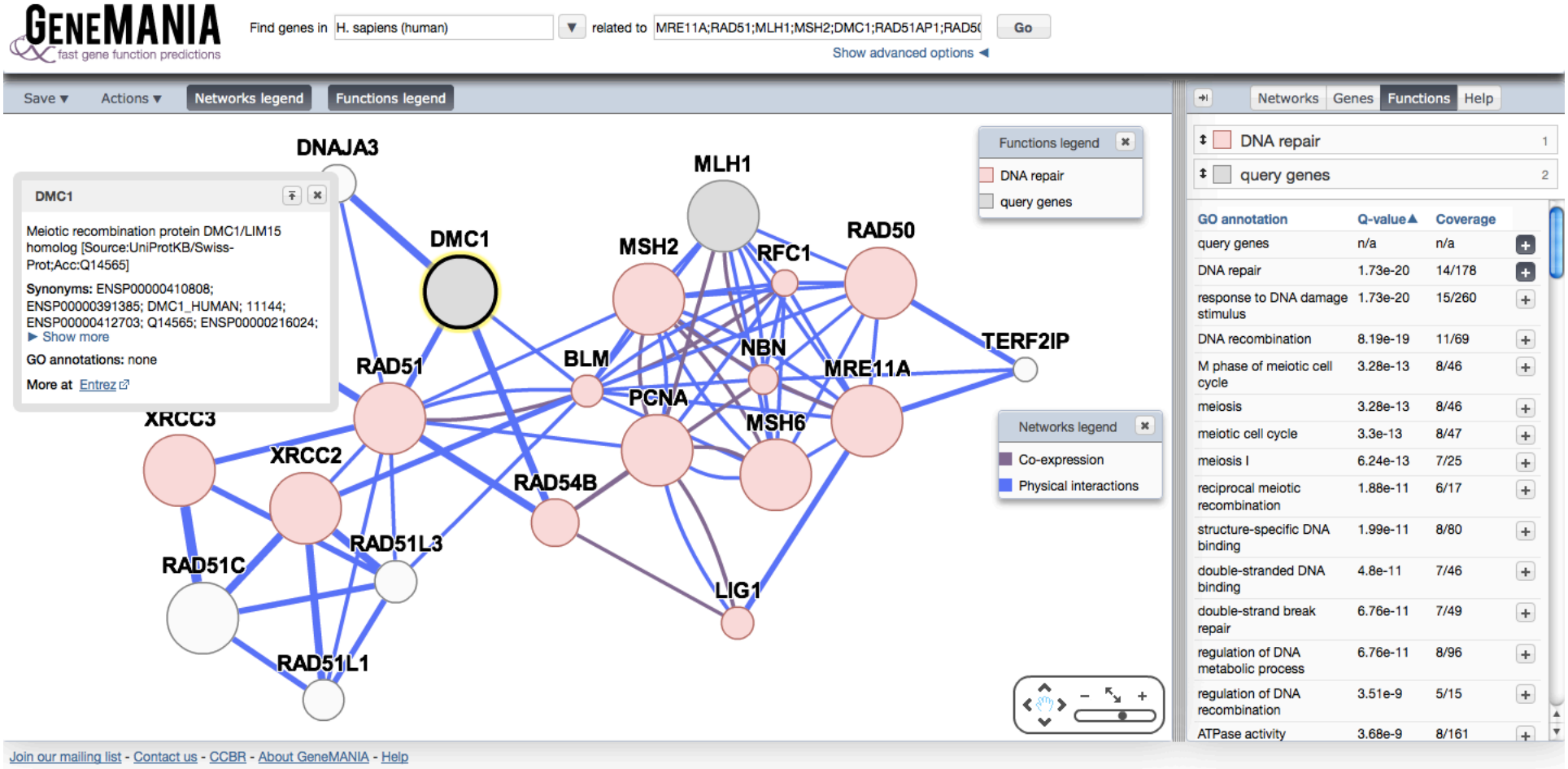
Cytoscape.js: HTML5 – iPad



http://www.genemania.org

Mostafavi S et al. Genome Biol. 2008;9 Suppl 1:S4

Warde-Farley D et al. Nucleic Acids Res. 2010 Jul; 8:W214-20.



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

Gene Function Prediction

Quaid Morris (Donnelly), Sara Mostafavi Rashad Badrawi, Ovi Comes, Sylva Donaldson, Max Franz, Christian Lopes, Farzana Kazi, Jason Montojo, Harold Rodriguez, Khalid Zuberi

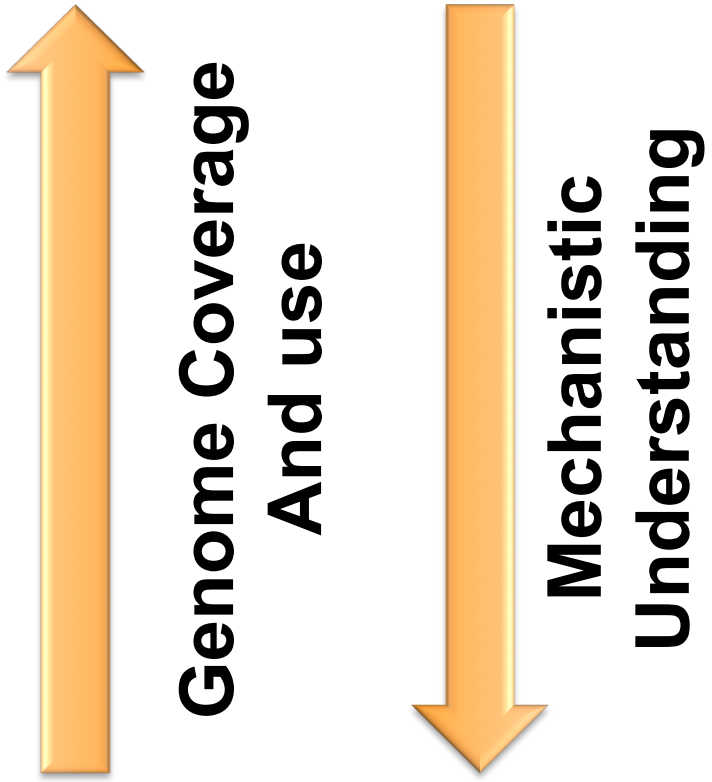
The Factoid Project

- Publishing in science
  - Highly inefficient
  - Outdated technology, difficult to search and compute
- http://www.elseviergrandchallenge.com/
  - Winner: http://reflect.ws/
- Pathway and network information database curation
  - Highly inefficient
- The factoid project

Max Franz, Igor Rodchenkov, Ozgun Babur, Emek Demir, Chris Sander

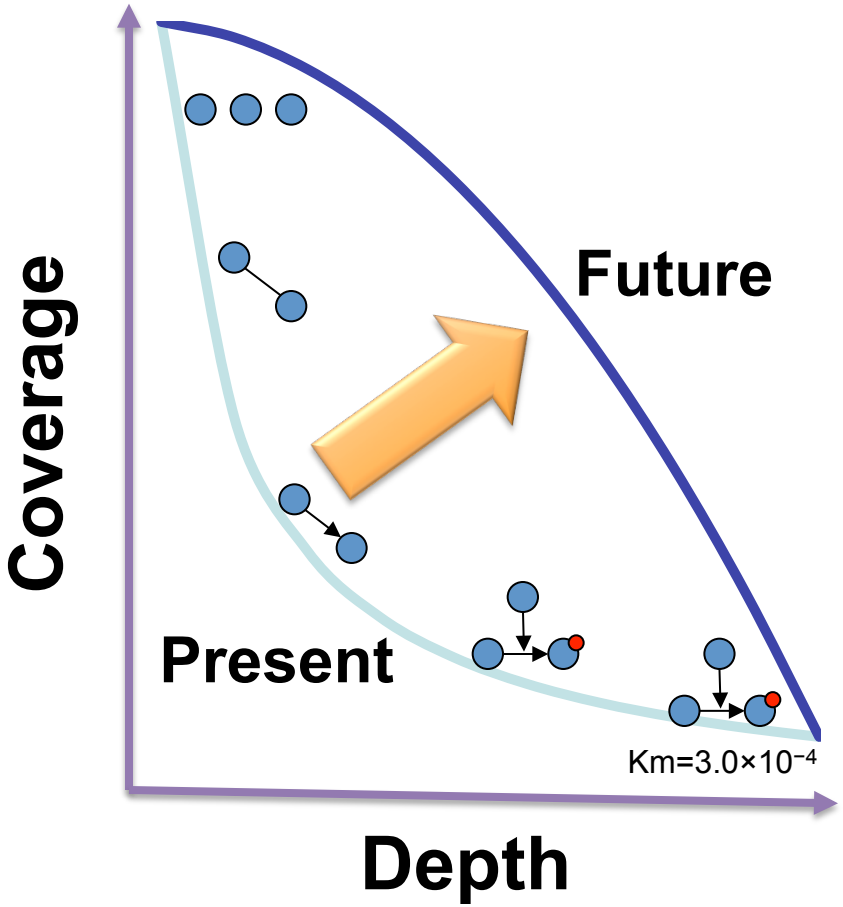
Pathway and Network Analysis

1. Gene set: pathway enrichment analysis
2. Network: network regions (modules), regulation
3. Process model: classical pathways
4. Simulation model: detailed models



Increase Coverage and Depth

Data and analysis methods



Cellular Process Representation

- Gene set
- Network
- Process model
- Simulation model

Depth

- Analysis methods need to keep up

Acknowledgements

Bader Lab
Domain Interaction Team
Chris Tan
Shirley Hui
Shobhit Jain
Brian Law
Jüri Reimand

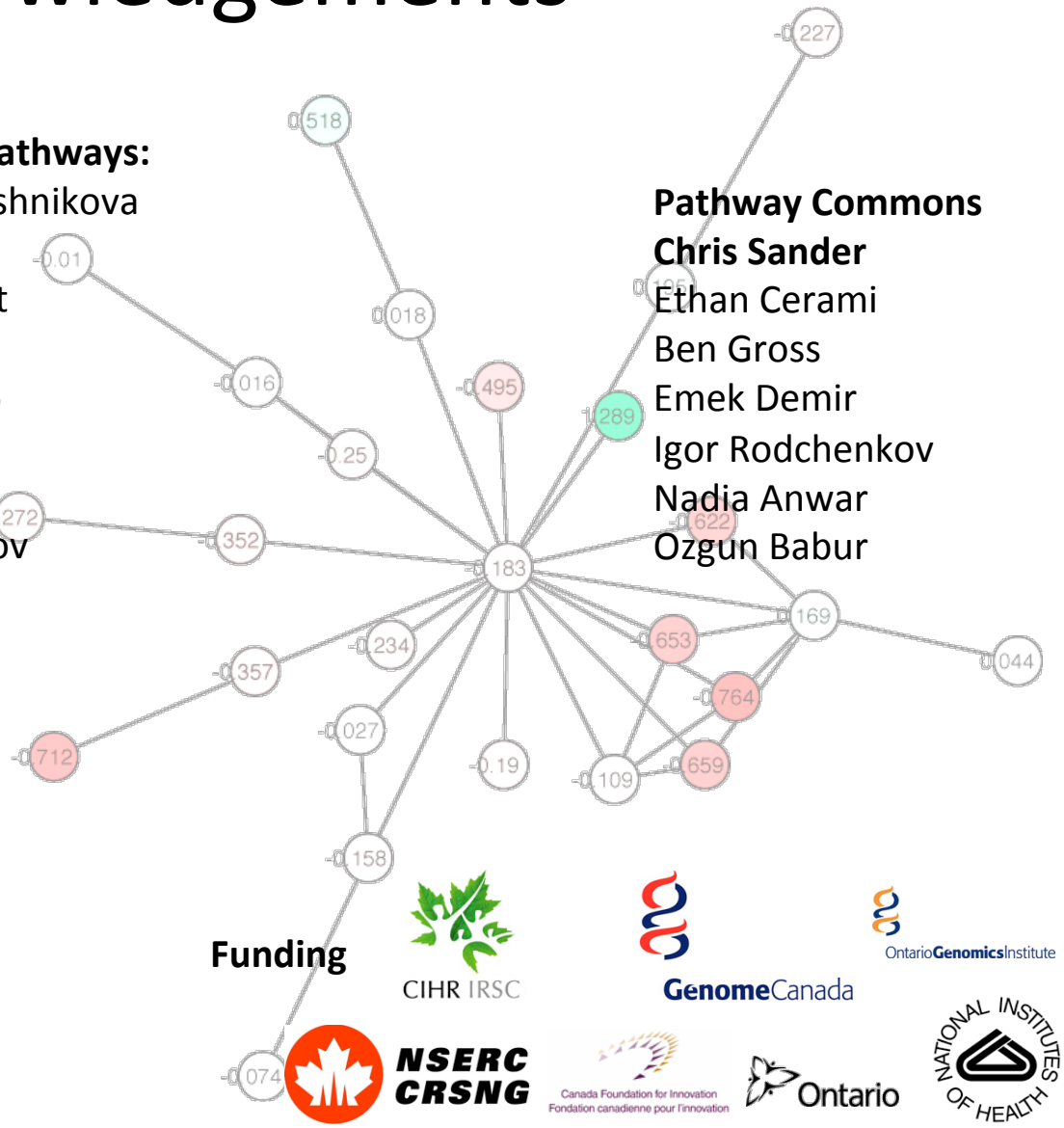
Former:
David Gfeller
Xiaojian Shao

Genetic Intx, Pathways:
Anastasia Baryshnikova
Iain Wallace
Magali Michaut
Ron Ammar
Daniele Merico
Ruth Isserlin
Vuk Pavlovic
Igor Rodchenkov

Pathway Commons
Chris Sander
Ethan Cerami
Ben Gross
Emek Demir
Igor Rodchenkov
Nadia Anwar
Ozgun Babur

www.GeneMANIA.org
Quaid Morris (Donnelly)
Rashad Badrawi, Ovi Comes, Sylva Donaldson, Christian Lopes, Farzana Kazi, Jason Montojo, Sara Mostafavi, Harold Rodriguez, Khalid Zuberi

Funding



http://baderlab.org