

Bridging data and modeling communities

**Ontologies of cellular networks, pathway
databases and pathway data retrieval tools**

Oliver Ruebenacker

Center for Cell Analysis and Modeling
University of Connecticut Health Center



**Second q-bio Conference
August 6st, 2008**



Outline

Communities and Data Online Services

- Pathways and Models
 - Knowledge bases and Simulators
 - XML versus OWL/RDF
 - SBML and BioPAX
- Pathway and model databases
 - Bunch of them
 - Modelling with VCell
 - BioPAX integration with Sybil

Two worlds

Modeling Community

- Data
- Do the math so that it can be calculated
- Data models
- Quantities
- Simulate and fit

Pathway Community

- Knowledge
- Link information so that it can be found
- Ontologies
- Relationships
- Reason and query

Two worlds – two descriptions

SBML

- “System Biological Markup Language”
- Quantitative information
- XML
- No term hierarchy
- “Designed by modelers”
- <http://www.sbml.org/>

BioPAX

- “Biological Pathway Exchange”
- Qualitative, no kinetics
- RDF / OWL
- Hierarchy of Terms
- “Designed by librarians”
- <http://www.biopax.org/>

Two worlds – two descriptions

SBML

- Units
- Compartments
- Parameters
- SpeciesTypes
- Species
- Reactions
 - Reactants
 - speciesReference
 - Products
 - speciesReference
 - KineticLaw

BioPAX

- Physical Entity
 - Protein, complex, small molecule, DNA, RNA
- Interaction
 - Conversion, control
 - Transport, catalysis, biochemical reaction, modulation
- Physical entity participant
 - Sequence participant

Notable Bio-Med Ontologies

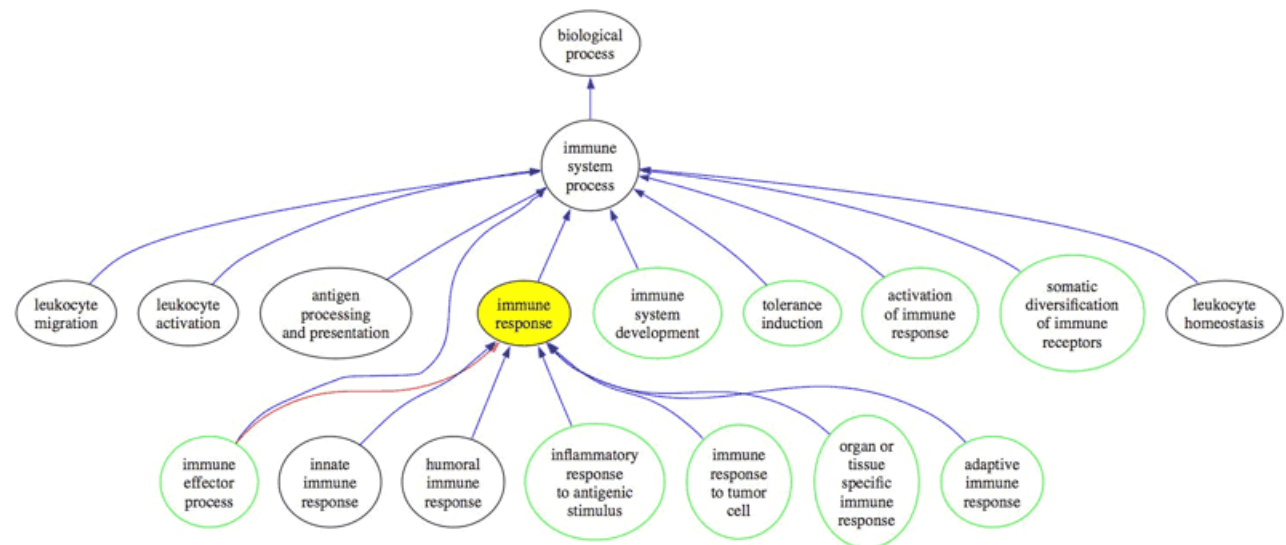
SBO = Systems Biology Ontology

BFO = Basic Formal Ontology

BioPAX = Biological Pathway Exchange

GO = Gene Ontology

RO (OBO) = Relation Ontology (Open Biomedical Ontologies)



Markup Languages

HTML = Hypertext Markup Language

XML = X-tensible Markup Language

SBML = Systems Biology Markup Language

```
<sbml>
```

```
  <ListOfReactions>
```

```
    <Reaction id="R1">
```

```
      ... stoichiometry="3" ...
```

```
    </Reaction>
```

```
  ...
```

```
  </ListOfReactions>
```

```
</sbml>
```

Resource Description Framework (RDF)

Resource = Something identifiable

- **Can have URI (sometimes looks like URL)**

Literal = piece of data (number, text, ...)

Node = Resource or Literal

Property = Resource describing a relationship

- **Object property: Between two Resources**
- **Datatype property: Resource and data**

Statement (or triple) = Subject (Resource) + Predicate (Property) + Object (Node)

RDF Example: BioPAX

```
mydb:EGF    rdf:type    bp:protein
bp:protein  owl:subClassOf  bp:physicalEntity
mydb:EGFR_bind  rdf:type    bp:bioch.Reac
mydb:EGFR_bind  bp:LEFT    mydb:EGFR_part1
mydb:EGFR_bind  bp:RIGHT    ...
mydb:EGFR_bind  bp:DELTA-G  ...
bp:bioch.Reac.  owl:subClassOf  bp:conversion
```

RDF Schema: RDF description of types and relationships used in RDF data

Web Ontology Language (OWL)

OWL can mean two things:

- (1) A usage of RDF, where restrictions on RDF properties are described by RDF statements**
- (2) A file format where RDF statements are encoded using XML**

In OWL, an RDF schema is called ontology

Biomodels.net

- **Database of models (in SBML)**
- **Minimal Information Requested In the Annotation of biochemical Models (MIRIAM)**
 - **A standard to annotate models**
- **Systems Biology Ontology (SBO)**
 - **Hierarchy of SB objects useful for annotation**
- **“Qualifiers”**
 - **Collection of relationships to be used in annotations**

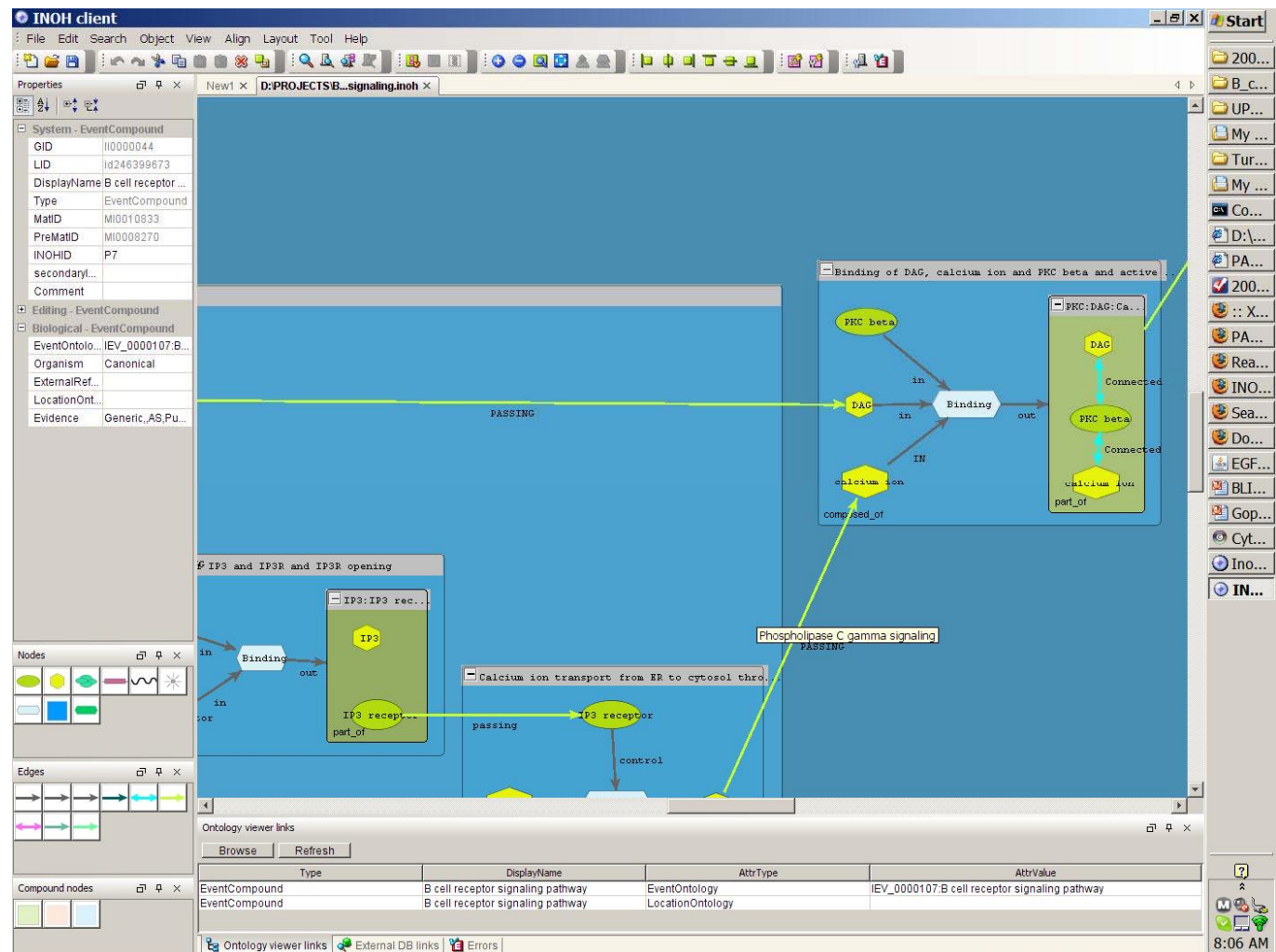
World of pathway databases

Data Source	Protein(SwissProt)	Coverage (SwissProt)	Interaction	Citation
Reactome	1229 (1194)	5% (8%)	21394	Vastrik <i>et al</i> , 2007
Panther	2997 (1670)	12% (12%)	75694	Mi <i>et al</i> , 2007
CellMap	567 (567)	2% (4%)	1195	cancer.cellmap.org
INOH	719 (711)	3% (5%)	11759	Kushida <i>et al</i> , 2006
NCI-Nature	593 (592)	2% (4%)	2900	pid.nci.nih.gov
NCI-BioCarta	936 (936)	4% (6%)	4752	pid.nci.nih.gov
KEGG	2033 (1947)	8% (13%)	11144	Kanehisa <i>et al</i> , 2004
Total	5283 (3847)	21% (27%)	118867	

- Majority supports BioPAX export: Reactome, CellMap, INOH, NCI-Nature, NCI-BioCarta
- Some support SBML export: Reactome, Panther. However, SBML export is limited.

INOH - Integrating Network Objects with Hierarchies

- <http://www.inoh.org/>
- Maintained by The University of Tokyo (Toshihisa Takagi, PI), Computational Biology Research Center (CBRC) (Ken Ichiro Fukuda, co-PI)
- Installable viewer,
- downloadable data,
- supports BioPAX



Reactome

- <http://reactome.org>
- Maintained by CSHL, EBI, NYU, and The GO Consortium
- Web viewer, API for export data, data download.
- Exports BioPAX, SBML

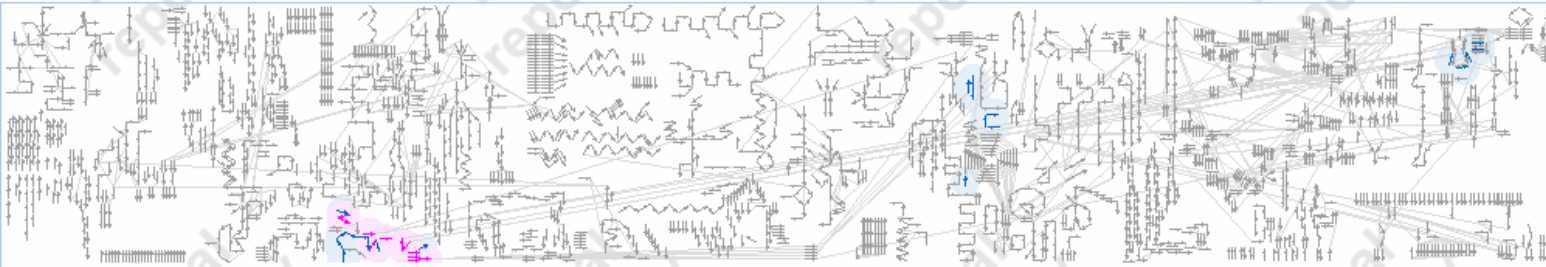
This is the Reactome internal repository. It includes data which have not been reviewed and released and is possibly incomplete. Our released data are at www.reactome.org.

[Home](#)
[About](#)
[Content](#)
[Documentation](#)
[Tools](#)
[Download](#)
[Help](#)
[Announcements](#)

Find everything with the EXACT PHRASE ONLY in Homo sapiens

Reactome - a curated knowledgebase of biological pathways

The data displayed is for Homo sapiens Use the menu to change the species. Check for cross-species comparison.

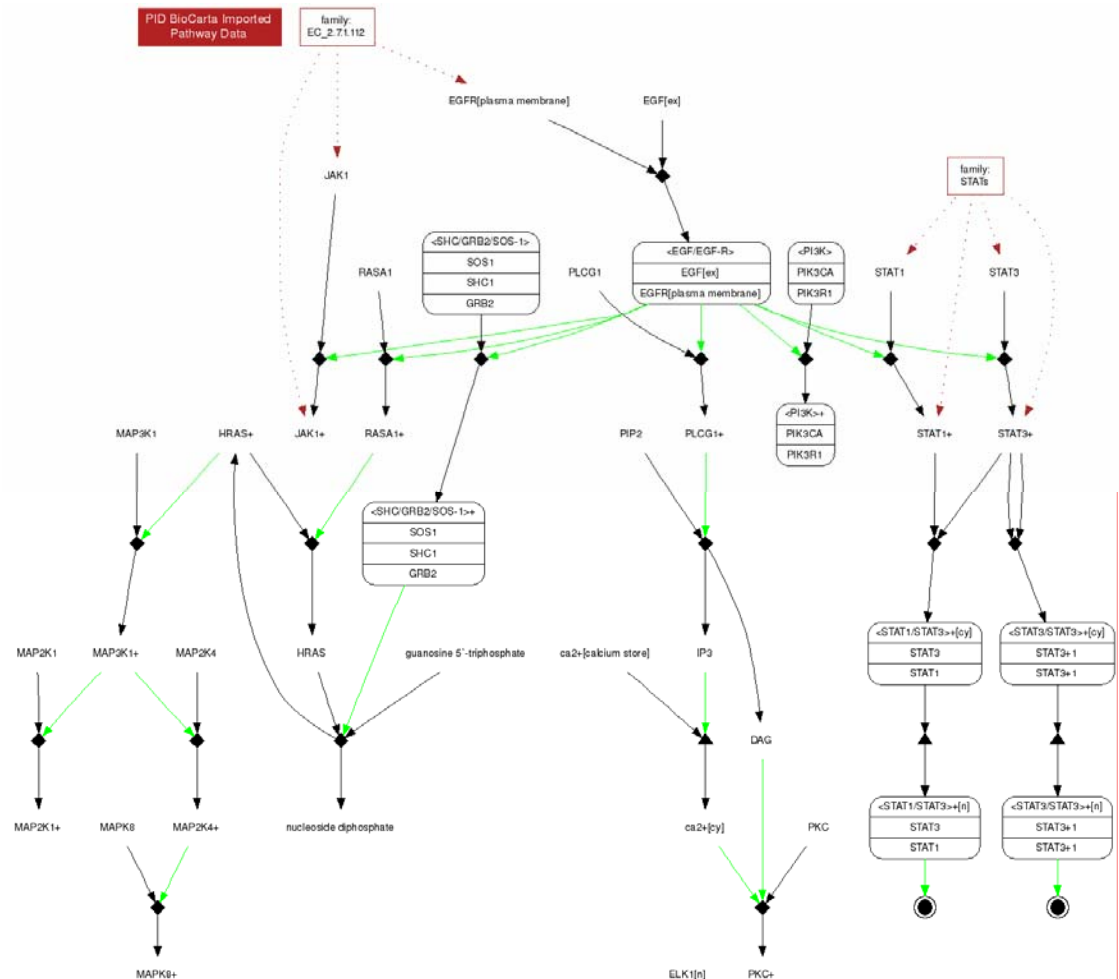


Reaction → | Experimentally confirmed reaction → | Manually inferred reaction → | Electronically inferred reaction → | Linked reactions →

Apoptosis	Bacterial virulence factors	Cell Cycle, Mitotic	Cell Cycle Checkpoints
Chromosome Maintenance	DNA Repair	DNA Replication	Electron Transport Chain
Epigenetic pathways	Gap junction trafficking and regulation	Gene Expression	Hematopoiesis
Hemostasis	HIV Infection	Influenza Infection	Integration of energy metabolism
Lipid and lipoprotein metabolism	Membrane Trafficking	Metabolism of amino acids	Metabolism of carbohydrates
Metabolism of non-coding RNA	Metabolism of vitamins and cofactors	Metabolism of xenobiotics	mRNA Processing
Metabolism of Nitric oxide	Nucleotide metabolism	Porphyrin metabolism	Post-translational protein modification
Pyruvate metabolism and TCA cycle	p53 tumor suppressor pathways	Regulation of beta-cell development	Regulatory RNA pathways
Signaling Pathways	Synaptic Transmission	Transcription	Translation

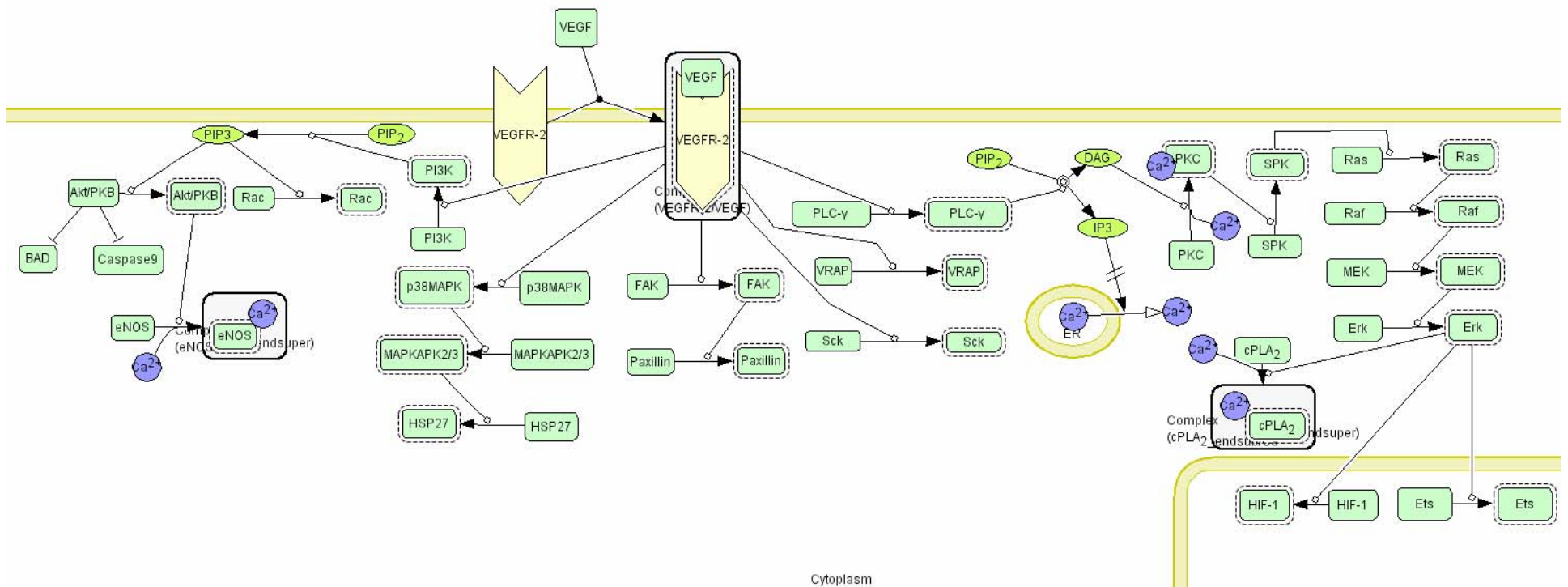
Pathway Interaction Database (PID)

- <http://pid.nci.nih.gov/>
- 66 Human Pathways 3799 Interactions Curated by NCI-Nature
- 306 Human Pathways 5833 Interactions Imported from [BioCarta/Reactome](#)
- Full download available
- Support BioPAX export



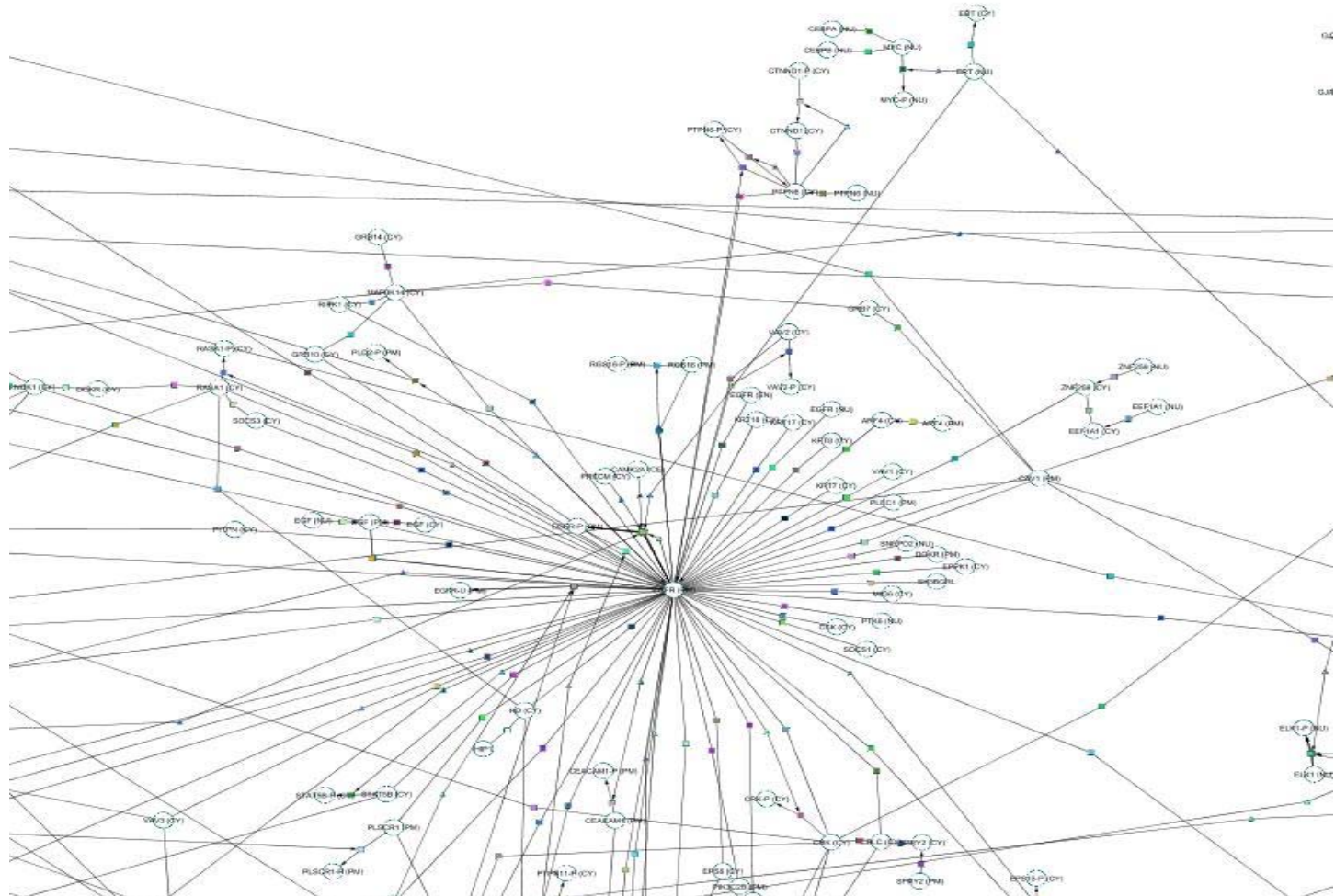
PANTHER (Protein ANalysis ThRUgh Evolutionary Relationships)

- <http://www.pantherdb.org/>
- Maintained by Evolutionary Systems Biology Group at SRI.
- Visualization using Cell Designer
- Exports to SBML, Based on GO



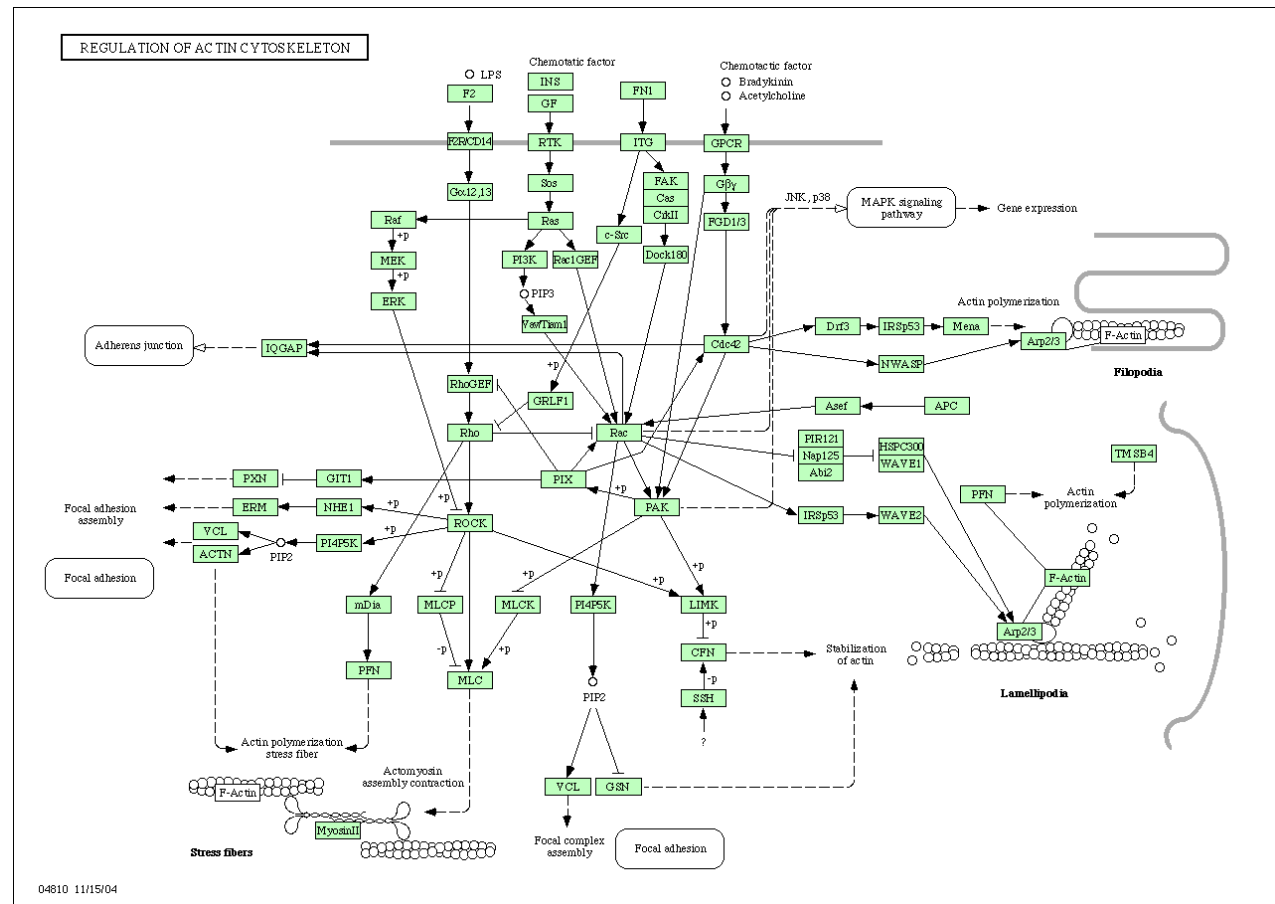
The Cancer Cell Map

- <http://cancer.cellmap.org/cellmap/>
- Maintained by Memorial Sloan-Kettering Cancer Center
- Provide API for access, Supports BioPAX, Visualized via Cytoscape



KEGG

- <http://www.genome.jp/kegg/pathway.html>
- Maintained by Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo
- Downloadable client, data on a server



Cytoscape Desktop (New Session)

File Edit View Select Layout Plugins Help

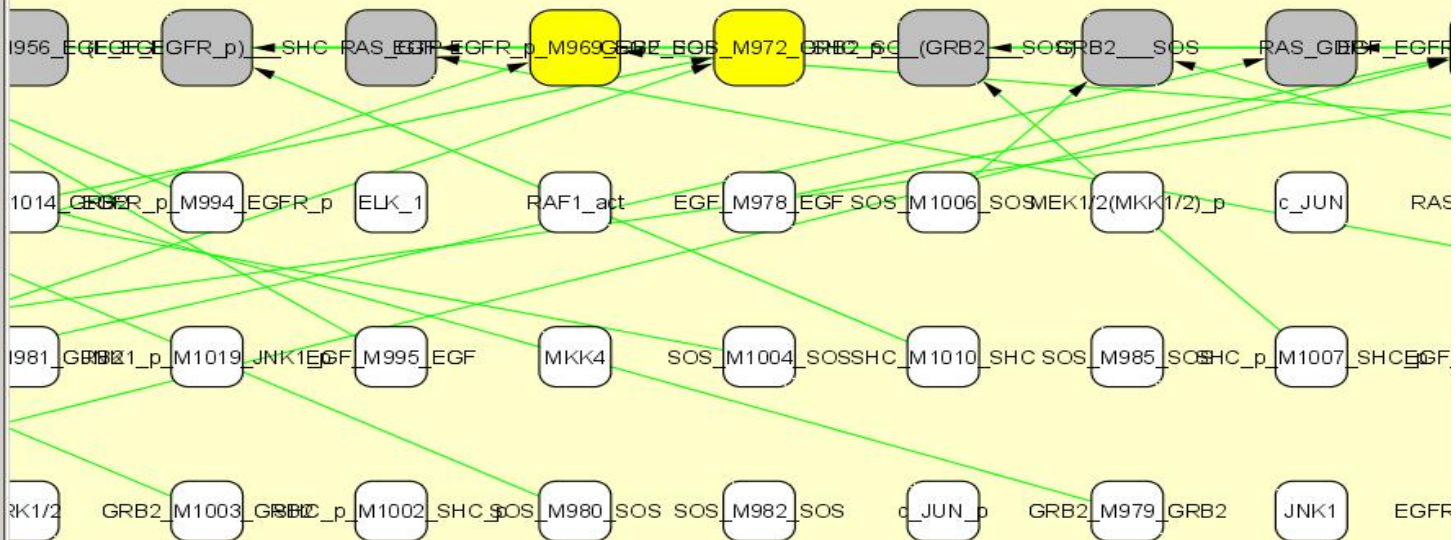
Control Panel

Network VizMapper™ Editor Filters

Search:

Network	Nodes	Edges
EGF RN	71(0)	56(0)
EGF PS	57(0)	56(0)
EGF PP	74(2)	33(0)

EGF PP



Data Panel



ID	canonicalName
EGF EGFR p M969 EGF EGFR p	EGF_EGFR_p_M969_EGF_EGFR_p
GRB2 SOS M972 GRB2 SOS	GRB2_SOS_M972_GRB2_SOS

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser

Virtual Cell (VCell)

- Simulator of cellular reaction networks
- Develop models with client, then run on servers
- Native format VCML (similar to SBML)
- Import and export to SBML
- Many functions: GUI model editor, compartments, , VFRAP, BioNetGen, ...
- <http://vcell.org/>

New VCell Features

Version 4.4

- Non-spatial stochastic modelling
- Field data (using images data as input to simulations)
- Annotations (MIRIAM compliant)
- Better SBML support

Version 4.5 (beta)

- Hybrid stochastic simulations
- Differential algebraic solver for non-spatial “fast” systems
- Math generation improvements
- Even better SBML support
- Formalized VCML as XML

Can BioPAX and SBML work together?

Conversions

- Binom (Cytoscape)
- Sbml2biopax (sbml.org)

Annotation

- Use BioPAX for annotations in SBML
- Add dummy elements to SBML to describe BioPAX features (dummy species)

Where BioPAX and SBML clash

- Sometimes an entity in BioPAX corresponds to multiple species types in SBML (e.g. A protein in varying phosphorylation states)
- Sometimes a species type in SBML corresponds to multiple entities in BioPAX (e.g. SNIPs, or generic participants in transport, translation and transcription)

Systems Biology Linker (Sybil)

- <http://vcell.org/biopax> (Yahogroups: sybil-users)
- Implemented:
 - BioPAX visualization and query
 - Conversion BioPAX -> SBPAX -> SBML
- Planned:
 - Model modules
 - Visualization: For SBML, improve for BioPAX
 - Editing/Merging multiple BioPAX files
 - Conversion SBML -> SBPAX -> BioPAX
 - API for automatic retrieval of BioPAX data
 - Implementing as a part of VCell

Systems Biology Pathway Exchange (SBPAX)

- SBPAX is an OWL-based ontology
- Used for conversion and mapping between SBML and BioPAX
 - Describes a common subset of SBML and BioPAX
 - Flexible enough to allow import from both
 - Distinction between an object and a model of the object

Acknowledgements

Virtual Cell

- Michael Blinov
- Ion Moraru
- Jim Schaff

BioPAX OBO Group

- Alan Ruttenberg
- Jonathan Reese
- Elgar Pichler
- Jeremy Zucker
- Joanne Luciano