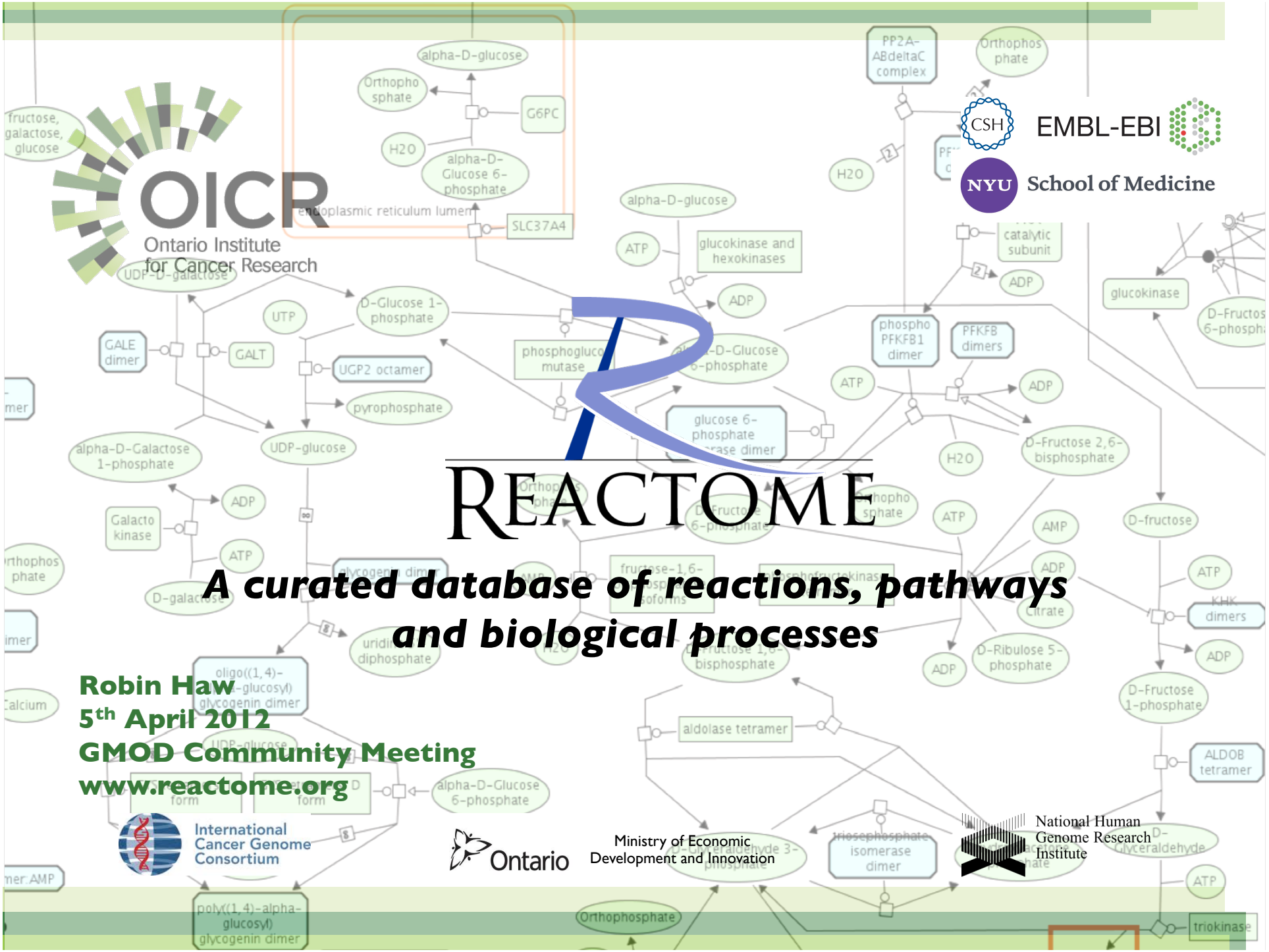
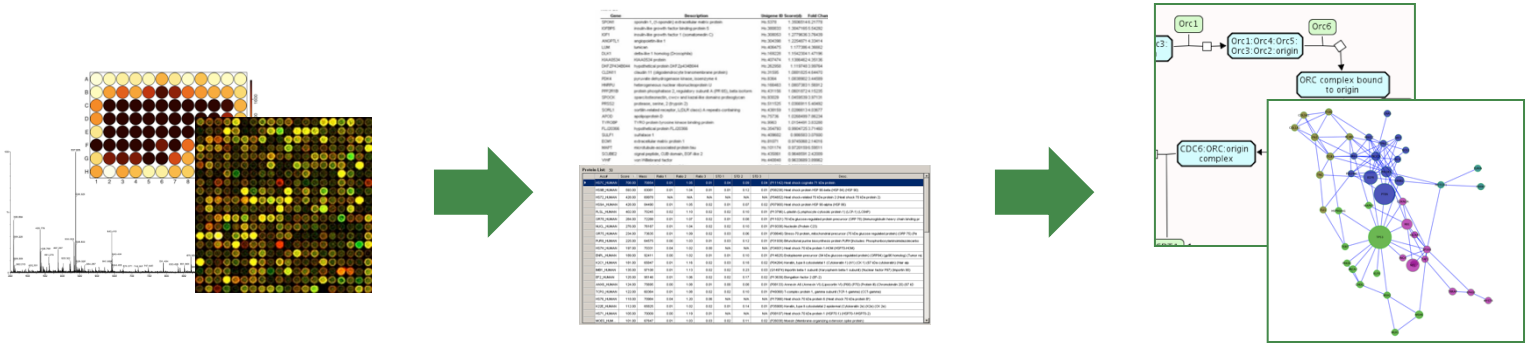

OICR
Ontario Institute
for Cancer Research
CSH
EMBL-EBI
NYU
School of Medicine
REACTOME
A curated database of reactions, pathways and biological processes
Robin Haw
5th April 2012
GMOD Community Meeting
www.reactome.org
International Cancer Genome Consortium
Ontario
Ministry of Economic Development and Innovation
National Human Genome Research Institute

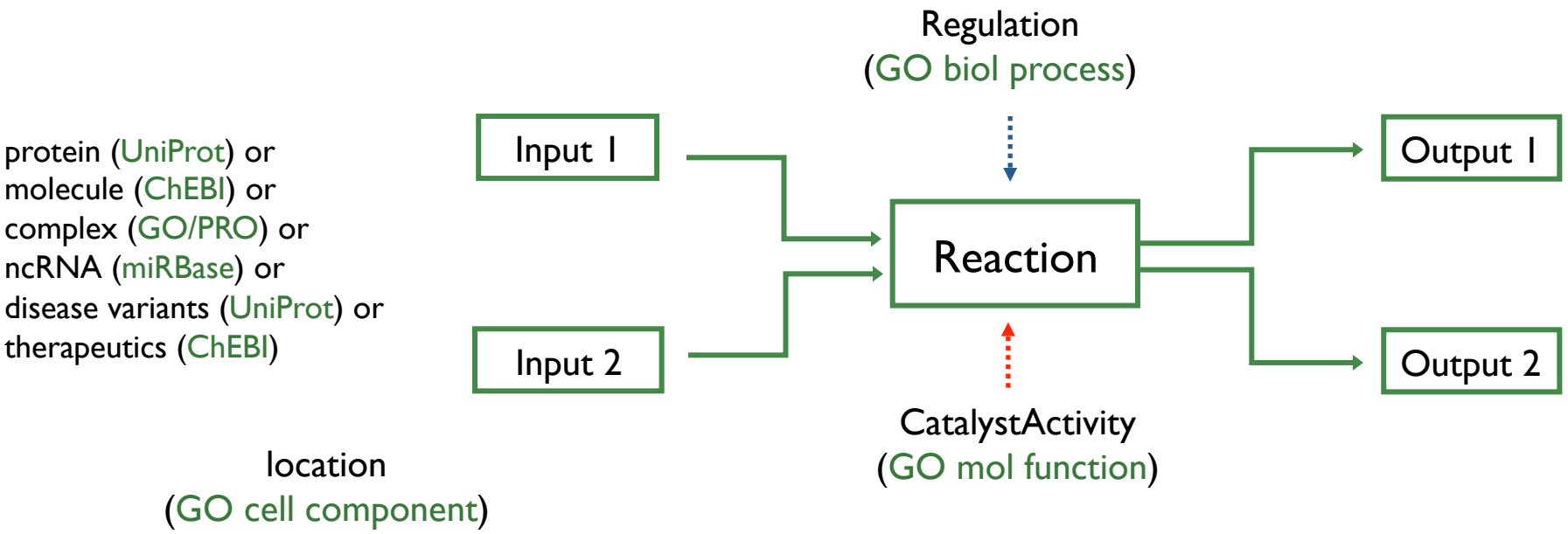
What is Reactome?

- Open source and open access pathway database
  - Metabolism, signaling, gene expression, DNA replication and repair and other biological processes in human biology
- Expert authored, manually curated and peer-reviewed
  - Rigorous curation standards – every reaction traceable to primary literature
  - Inferred and manually curated model organism pathways
- Extensively cross-referenced to external bioinformatics databases
- Provides tools and datasets for browsing, querying, analyzing and visualizing pathway data



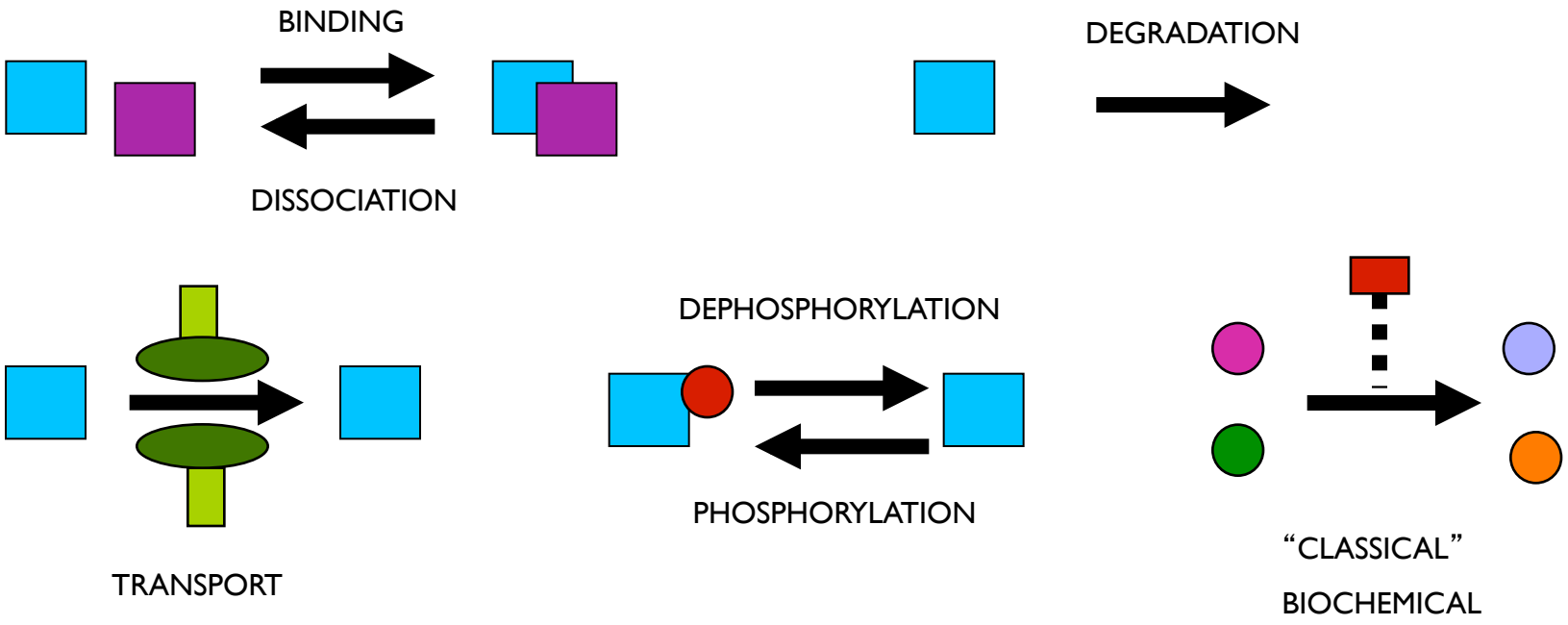
Data model in a nutshell

- Basic "unit" of Reactome

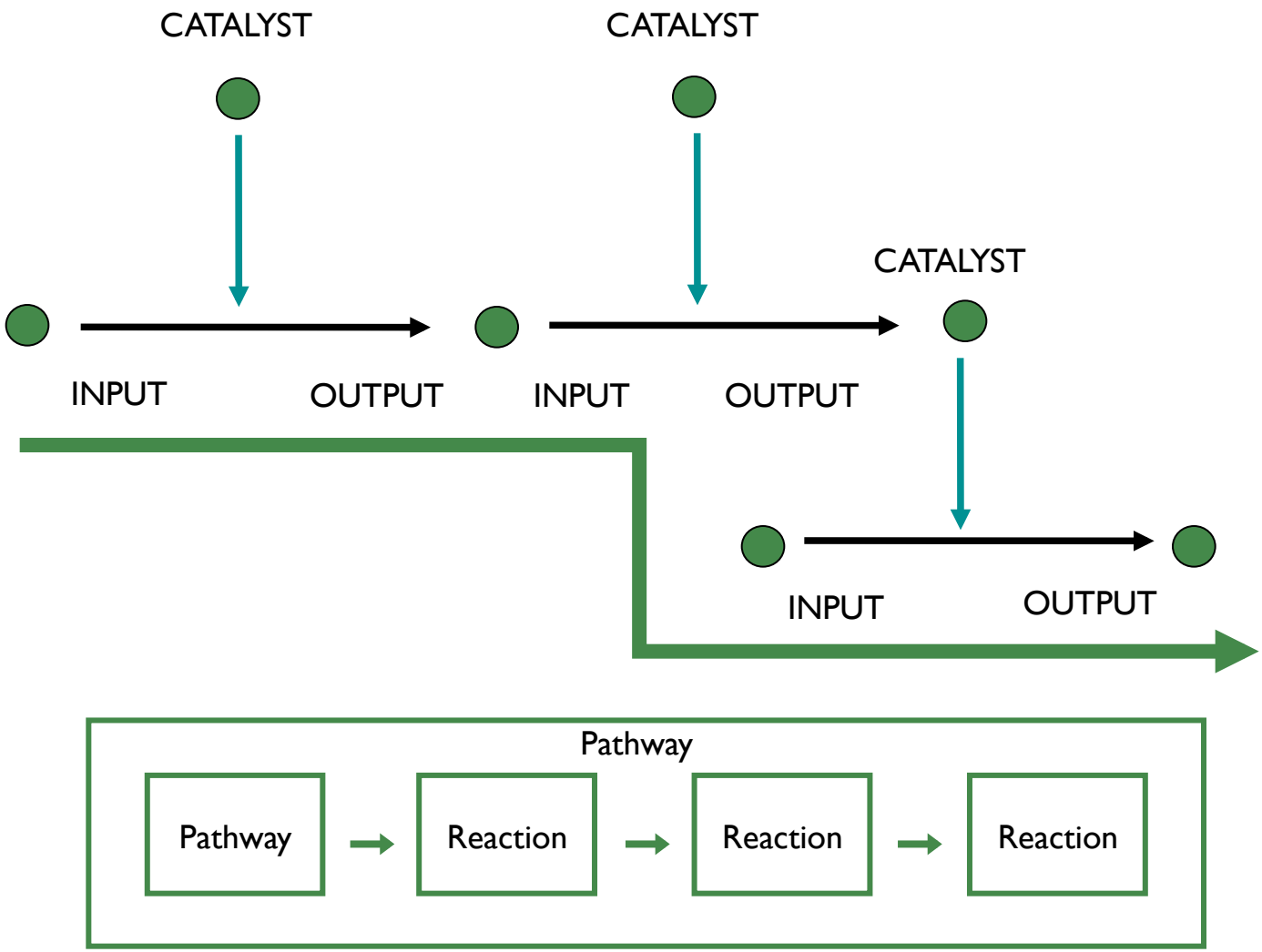


Theory - Reactions

- Represents many events and states found in biology.



Reactions Connect into Pathways



Where the Data Comes From?

- Recruit bench scientists to write modules.
  - All molecules are identified explicitly
  - All assertions backed by literature references
- Curator works with author to ensure consistency and completeness
- Module is checked by peer review and software prior to publication
- Public release of curated data every 3 months
- Regular pathway updates and rolling review

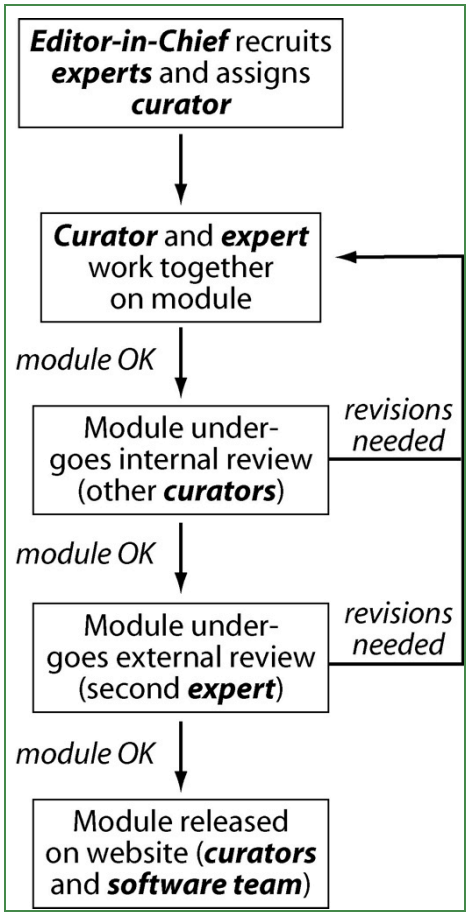


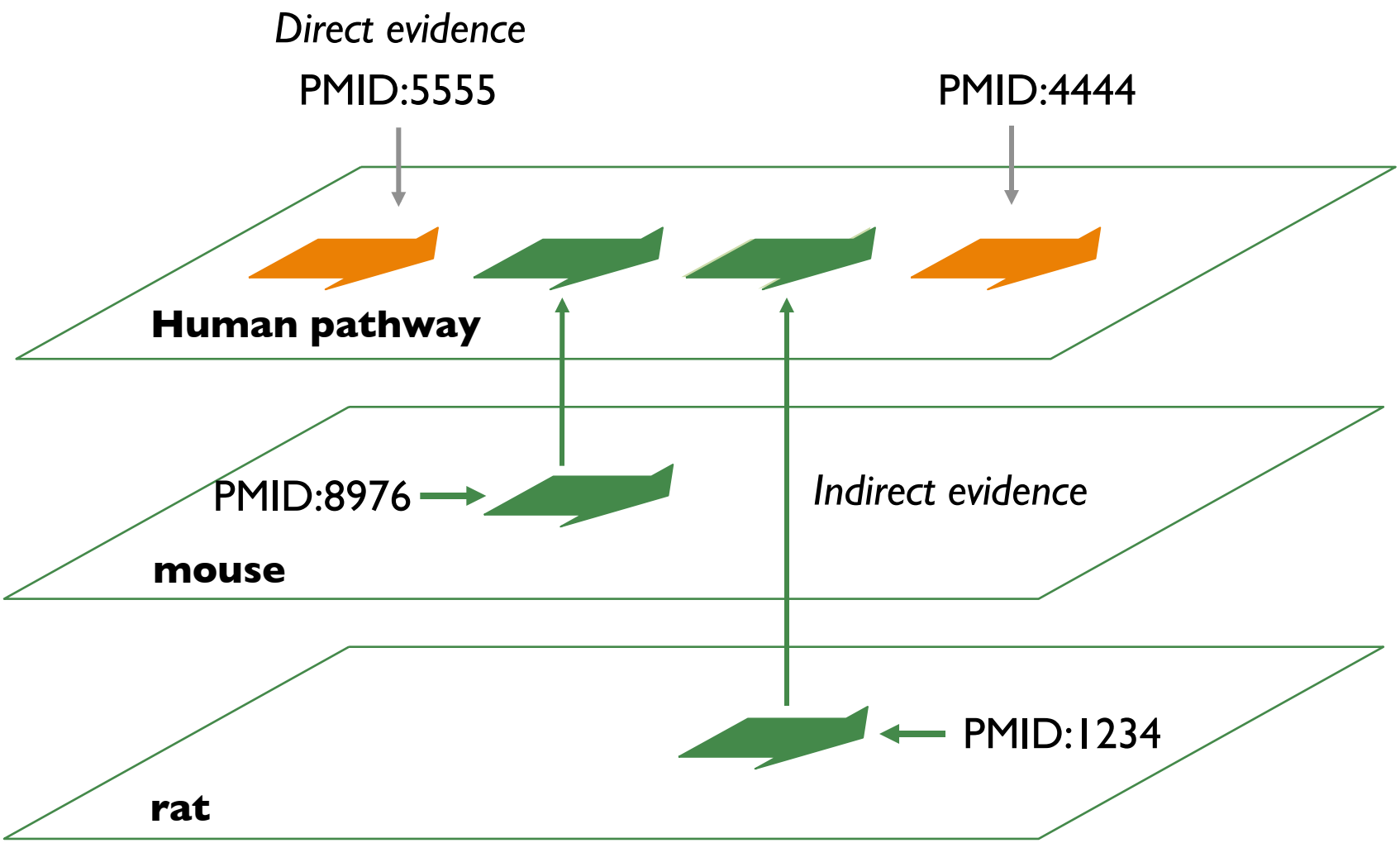
Table of Contents and Editorial Calendar

Topic	Authors	Released	Revised	Reviewers	Editors
Apoptosis [Homo sapiens] - Extrinsic Pathway for Apoptosis (DOI) - Intrinsic Pathway for Apoptosis (DOI) - Apoptotic execution phase - Regulation of Apoptosis	Alnemri, E, Hengartner, M, Tschopp, J, Tsujimoto, Y, Hardwick, JM, Gillespie, ME, Matthews, L, Matthews, L, Jakobi, R, Gopinathrao, G, Schulze-Osthoff, K, Garapati, P V, Ranganathan, S, Williams, MG	2004-09-20		Hengartner, M, Ranganathan, S, Vaux, DL, Chang, E, Widlak, P, Cooper, HM, Silverman, N, Lemaitre, B	Gopinathrao, G, Matthews, L, Gillespie, ME, Joshi-Tope, G, Matthews, L, Garapati, P V
Cell Cycle [Homo sapiens] - Cell Cycle Checkpoints - Cell Cycle, Mitotic - Chromosome Maintenance	Hoffmann, I, Khanna, KK, O'Connell, M, Walworth, N, Yen, TJ, Bosco, G, Matthews, L, Orlic-Milacic, M, Gillespie, ME, Yen, T, Blackburn, EH, Seidel, J, D'Eustachio, P, May, B, Borowiec, JA, Pagano, M, Lorca, T, Castro, A, Matthews, L, Gopinathrao, G, Tom, S, Bambara, RA, Lee, KS, Davey, MJ, O'Donnell, M, Tye, BK, Sanchez, Y, Joshi-Tope, G, Watanabe, N, Hunter, T	2011-12-06		Sanchez, Y, Knudsen, E, Hardwick, KG, Manfredi, J, MacPherson, D, Grana, X, Price, C, Bolcun-Filas, E,	Matthews, L, Matthews, L, Gopinathrao, G, Joshi-Tope, G, May, B, D'Eustachio, P, Gillespie,
Cell-Cell communication [Homo sapiens] - Cell junction organization (DOI) - Signal regulatory protein (SIRP) family interactions (DOI) - DSCAM interactions - Nephrin interactions (DOI)	Garapati, P V, de Bono, B, Matthews, L	2011-09-20			
Circadian Clock [Homo sapiens] (DOI) - BMAL1:CLOCK/NPAS2 Activates Circadian Expression - Circadian Repression of Expression by REV-ERBA - RORA Activates Circadian Expression	May, B	2010-12-14 UPDATED			

Next Release - Ver 41 : June 2012 Reactome

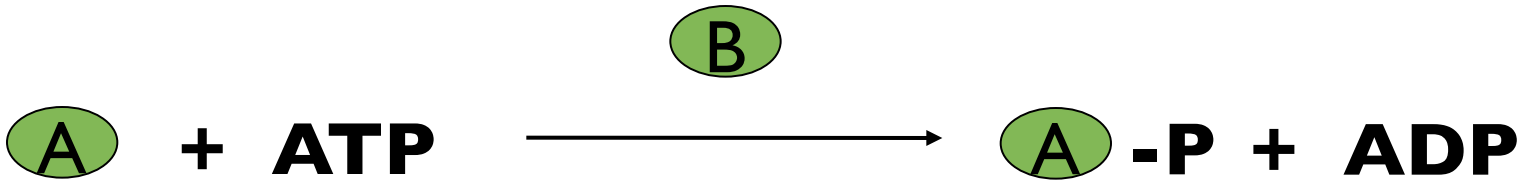
Curator	Pathway Topic	Author	Reviewer
Bijay Jassal	Activation of arylsufatases	B Jassal	R Stephan
Steve Jupe	Chromatin modifying enzymes	TBA	TBA
Phani Garapati	DAP12 and C-type Lectins in phagocytosis	P Garapati	TBA
Phani Garapati	FCGR mediated phagocytosis	P Garapati	TBA
Marc Gillespie	Fertilization	M Gillespie	TBA
Karen Rothfels	FGFR in disease	S Ezzat	TBA
Bruce May	GLUT4 Translocation	B May	A Klip
M Williams	Glycerophospholipid Biosynthesis Pathway	M Williams	TBA
Bijay Jassal	Glycosaminoglycan metabolism	B Jassal	TBA
Bruce May	Hydration of Carbon Dioxide	B May	TBA
Bijay Jassal	Latent infection of Homo sapiens with Mycobacterium tuberculosis	R Stephan	TBA
Bruce May	Metabolism of Angiotensinogen to Angiotensin	B May	TBA
Mark Williams	Metabolism of Arachidonic acid	M Williams	TBA
Phani Garapati	Metabolism Linoleic acid and alpha-linolenic acid	P Garapati	G Burdge
Mark Williams	Metabolism of phosphatidylinositol	M Williams	TBA
Mark Williams	Metabolism of inositol phosphate	M Williams	TBA
Phani Garapati	MHC Class II mediated antigen processing & presentation	P Garapati	TBA
Bruce May	Mitochondrial biogenesis	B May	TBA
Bruce May	Mitochondrial iron-sulfur cluster biogenesis	B May	TBA
Bruce May	O2/CO2 exchange	B May	TBA
Bijay Jassal	Post translational modifications	B Jassal	TBA
Bruce May	Regulation of cholesterol by SREBP's	B May	TBA
Bruce May	Regulation of hypoxia-inducible factor by oxygen	B May	TBA
Marija Milacic	Signaling by TGF beta (update)	TBA	TBA
Mark Williams	Ubiquinol Biosynthesis	M Williams	TBA

Evidence Tracking – Inferred Reactions

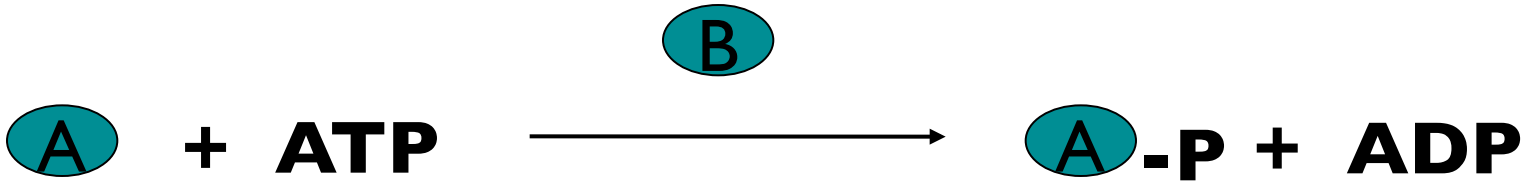


Data Expansion – Projecting to Other Species

Human



Mouse

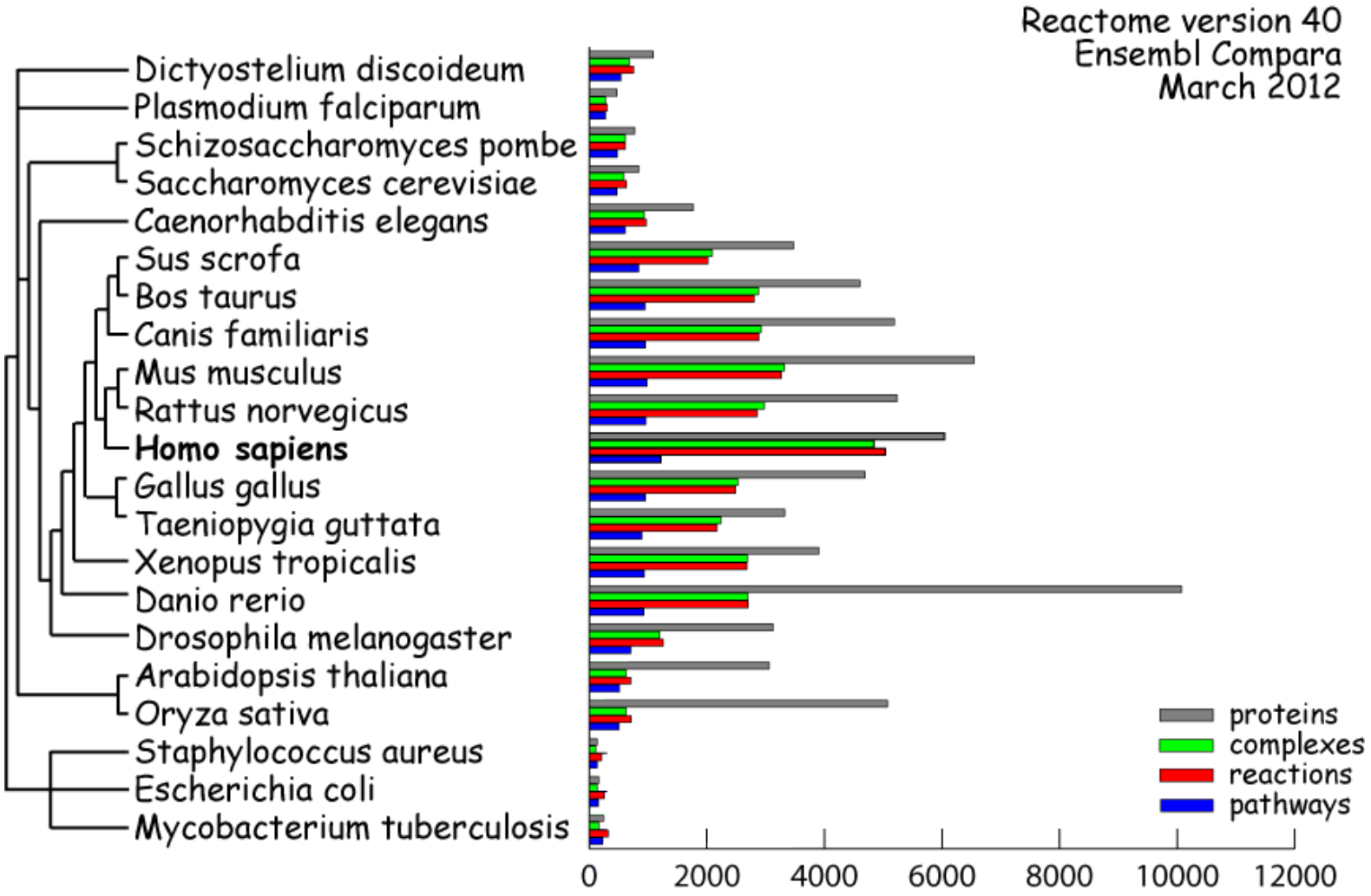


Drosophila

Reaction not inferred

No ortholog –
Protein not inferred

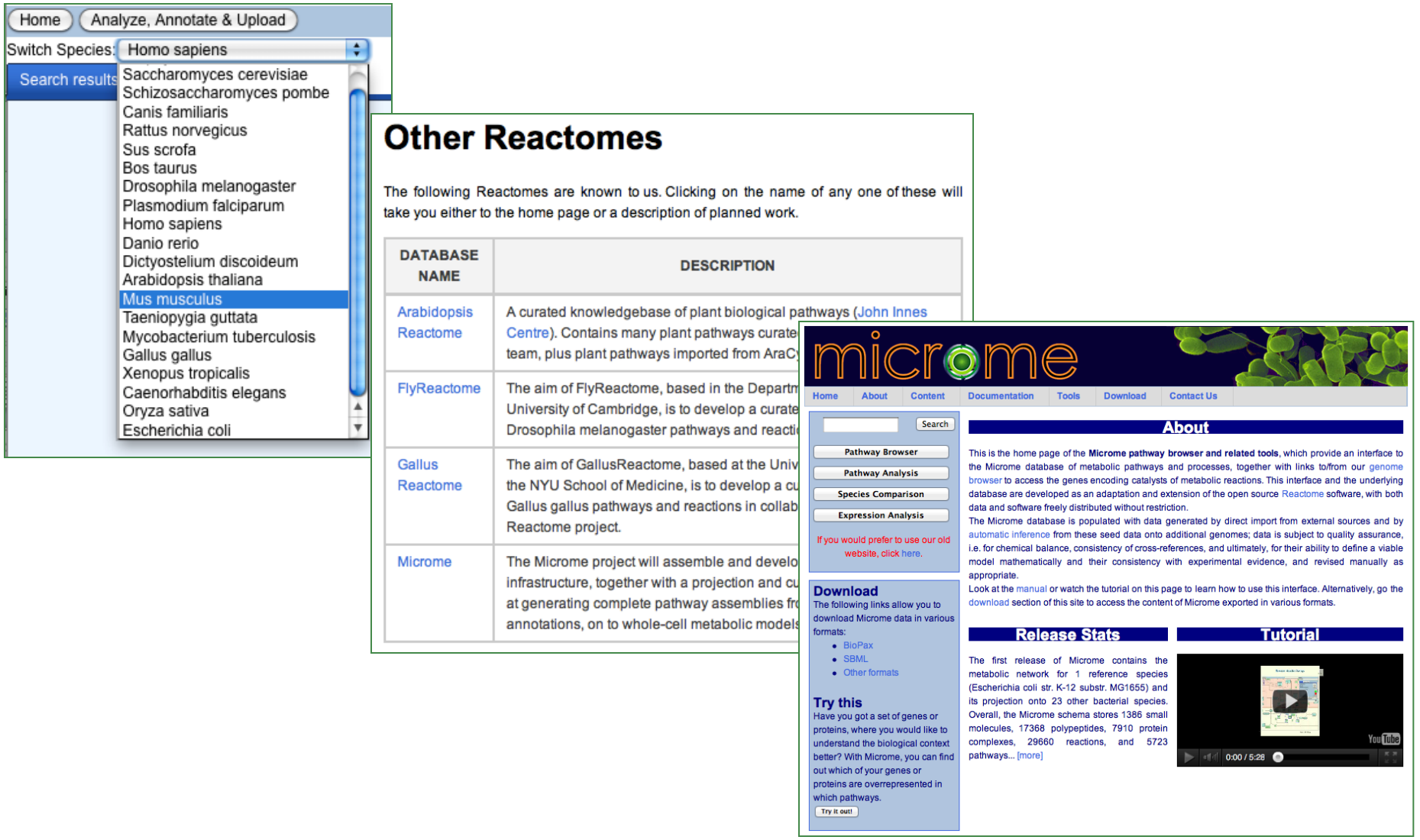
Data Expansion – Projecting to Other Species



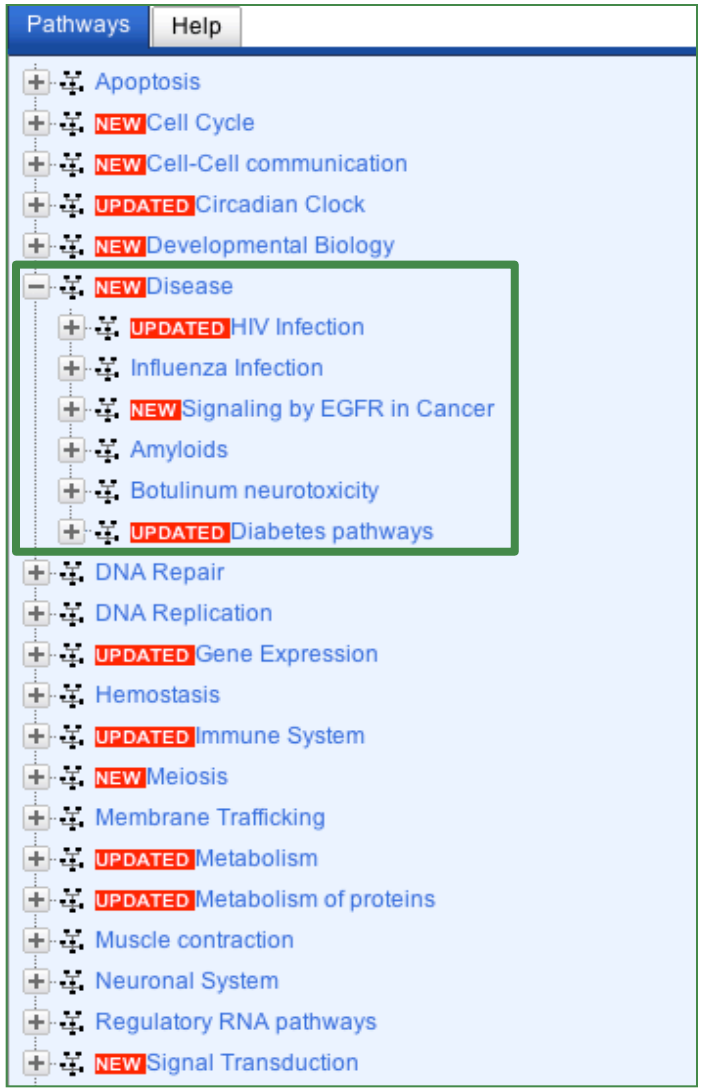
OICR Ontario Institute for Cancer Research

REACTOME

Inferred and Manually Curated M/O Reactome



Focus on disease areas

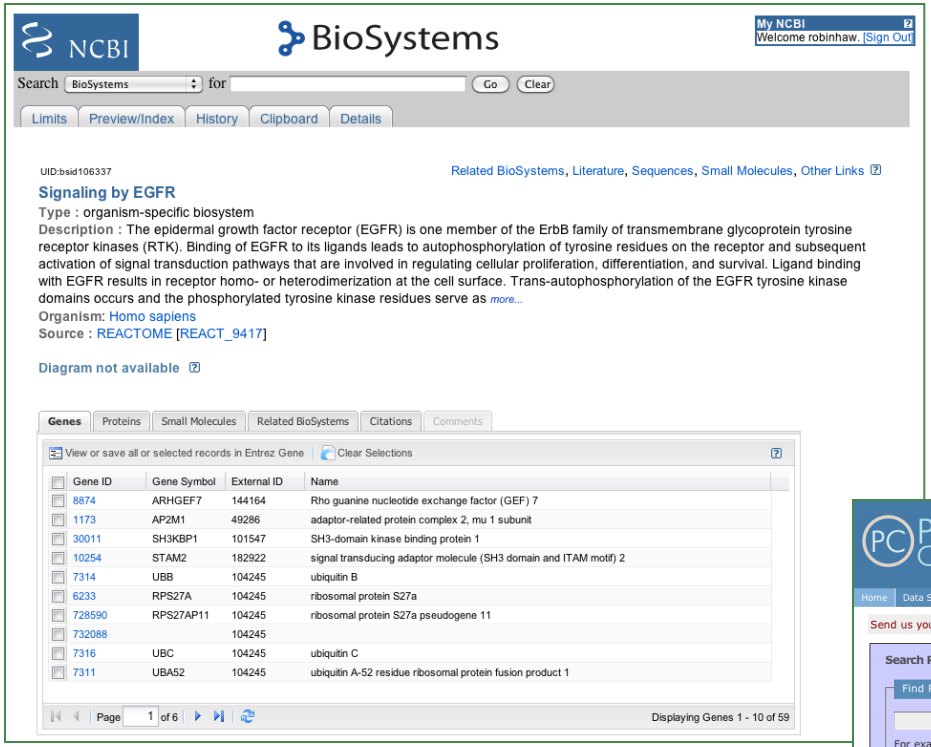


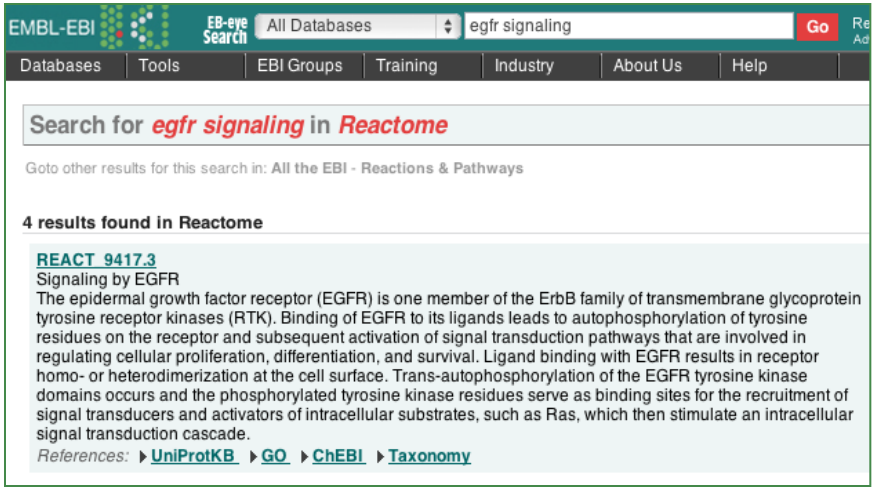
- Infectious Proteins: An infection introduces new proteins into the cell
- Cancer mutants: Precisely annotate protein sequence changes
- Gain/loss of function: Show how cancer mutations affect protein function
- Anti-cancer therapeutics: Visualize mode of action of anti-cancer drugs
- Pathology: List different cancer types in which a mutant protein has been identified

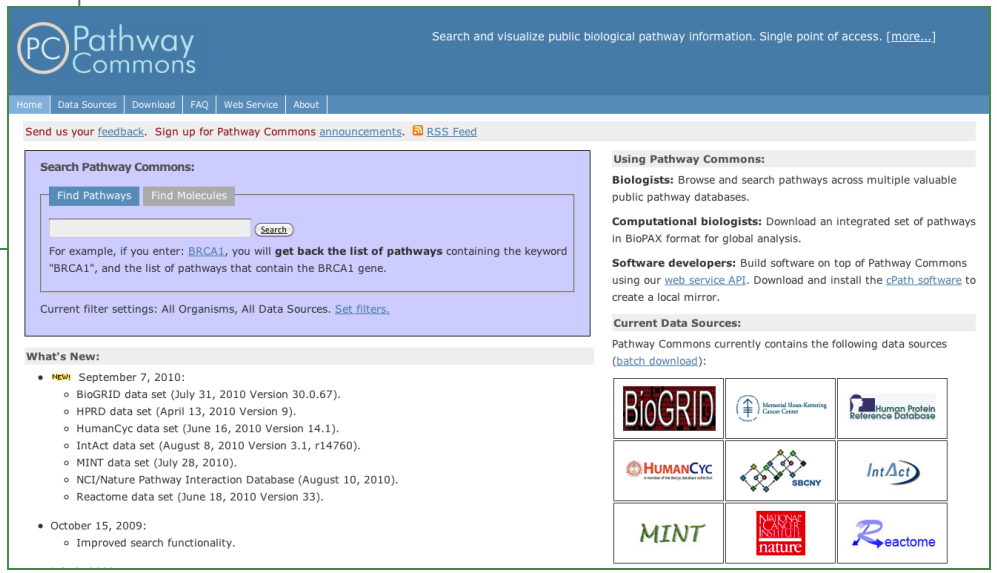
Link-outs from Reactome

- Gene Ontology - Molecular Function, Cellular Component, Biological process
- Small molecules - KEGG Compound, ChEBI, PubChem Compound
- Proteins - UniProt
- Genes and Genomes - Ensembl, Entrez Gene, dbSNP , KEGG Gene, RefSeq, HapMap, UCSC
- Disease - COSMIC, Disease Ontology
- Literature evidence - PubMed
- Structural Biology - PDB, DockBlaster
- Toxicology - Comparative Toxicogenomics Database
- Other - BioGPS

Data Warehouses integration and other exports







Other Data Exports

- Gene Ontology, Protein Ontology
- HapMap and UCSC Genome Browsers
- GSEA
- WikiPathways, Wormbase

OICR Ontario Institute for Cancer Research

REACTOME

Reactome Home Page

www.reactome.org

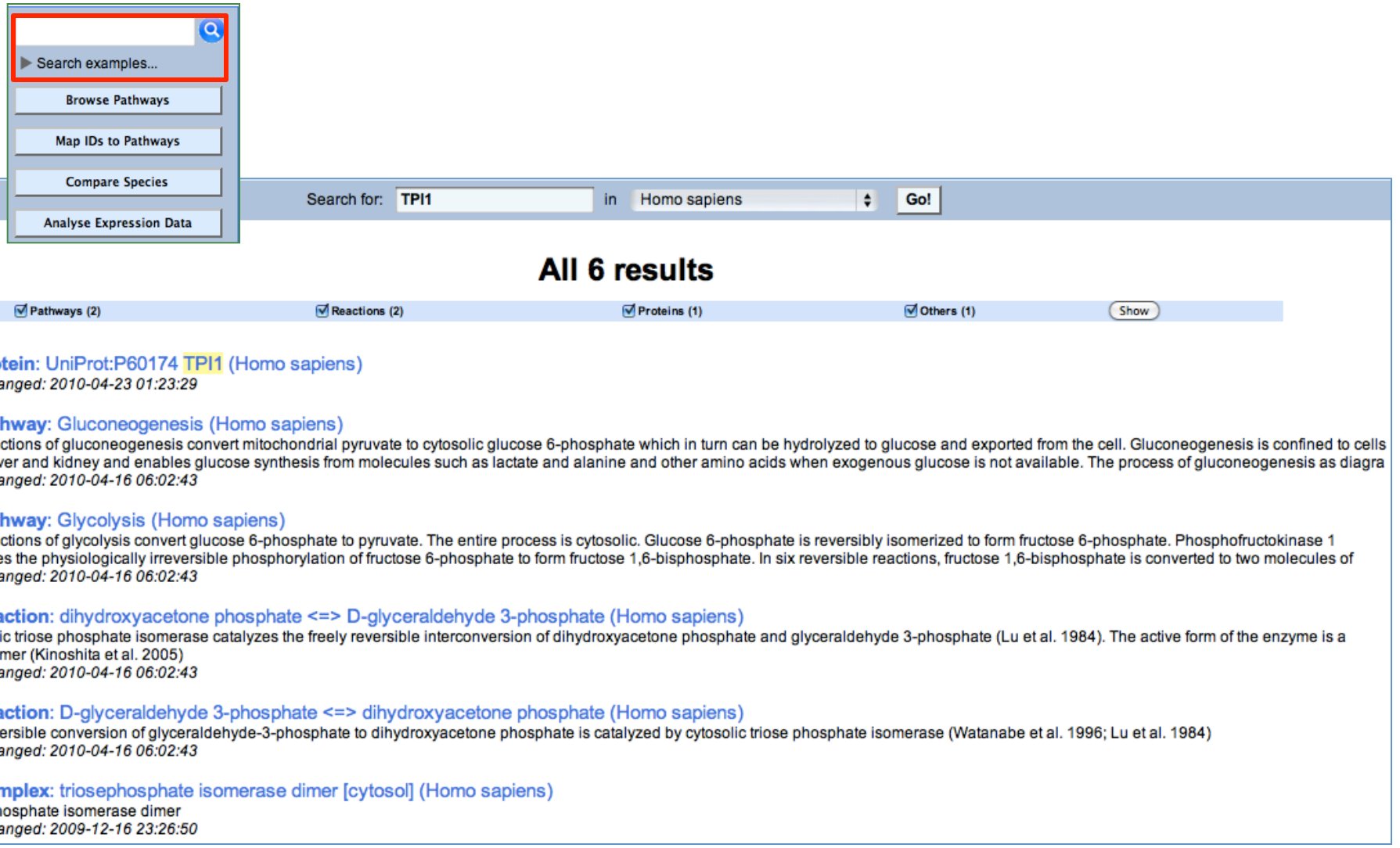
Documentation

User Guide: http://www.reactome.org/userguide/Usersguide.html

OICR Ontario Institute for Cancer Research

REACTOME

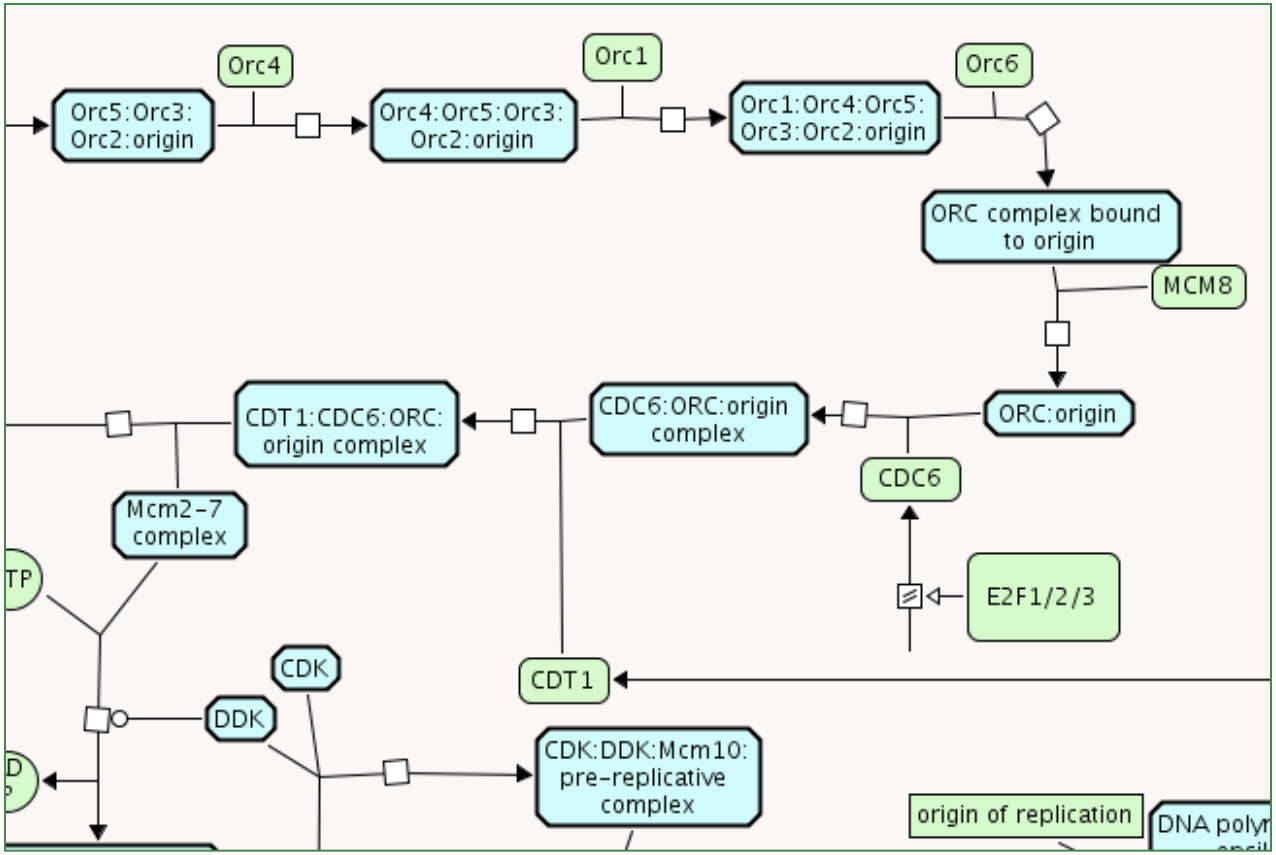
Basic Search



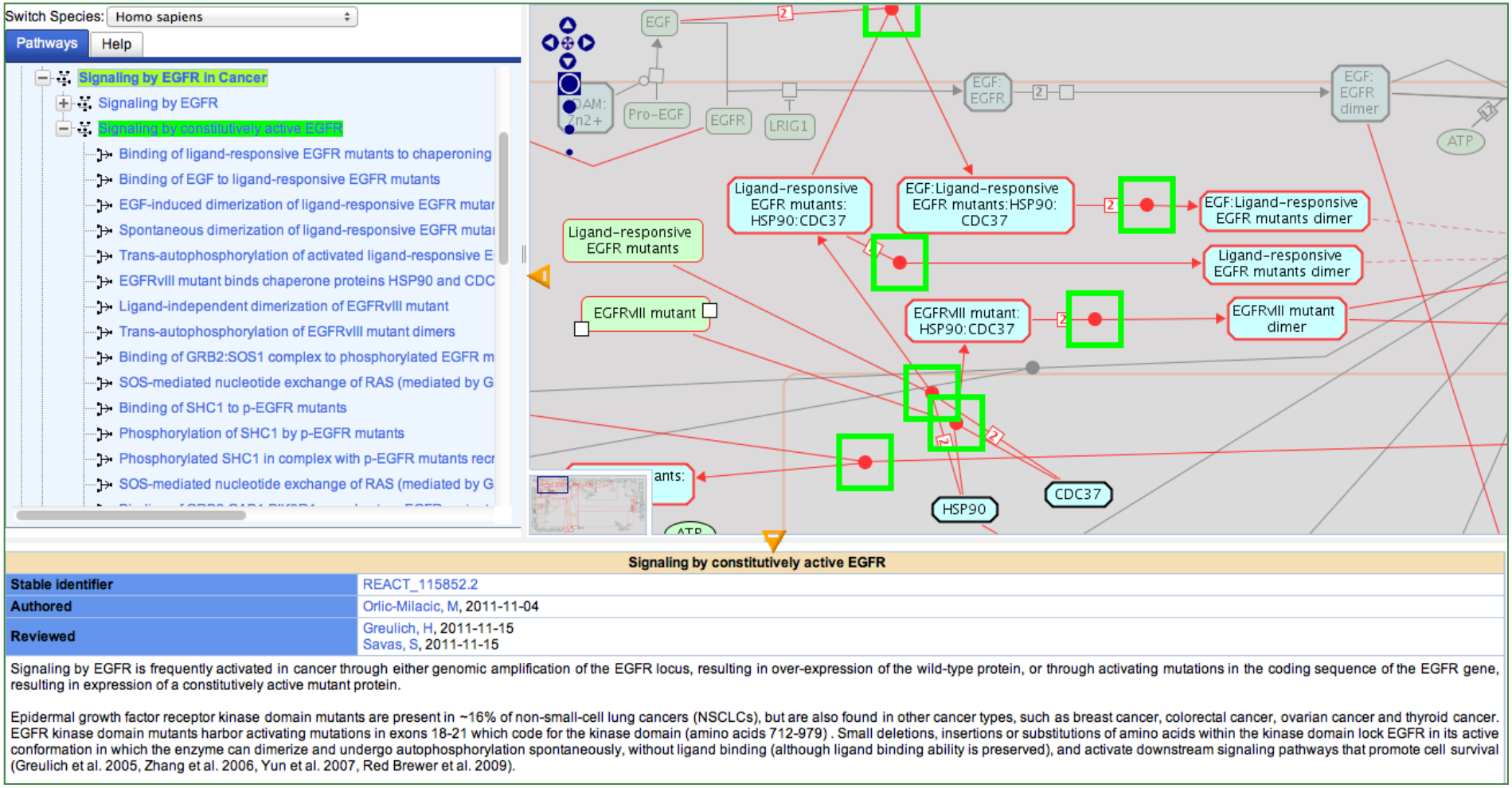
Divide reaction space into ~160 canonical pathways

- DNA Replication
  - Mitotic M-M/G1 phases
  - Synthesis of DNA
  - Regulation of DNA replication
- UPDATED Gene Expression
- Hemostasis
- HIV Infection
- UPDATED Immune System
- Influenza Infection
- Integration of energy metabolism
- Integrin cell surface interactions
- Interactions of the immunoglobulin su
- Meiotic Recombination
- Membrane Trafficking
- UPDATED Metabolism of amino acids
- Metabolism of carbohydrates
- NEW Metabolism of hormones
- Metabolism of lipids and lipoproteins
- Metabolism of nitric oxide
- Metabolism of nucleotides
- Metabolism of porphyrins
- Metabolism of proteins
- UPDATED Metabolism of RNA

Each represented by an SBGN-like diagram



Browsing EGFR Signaling Pathway



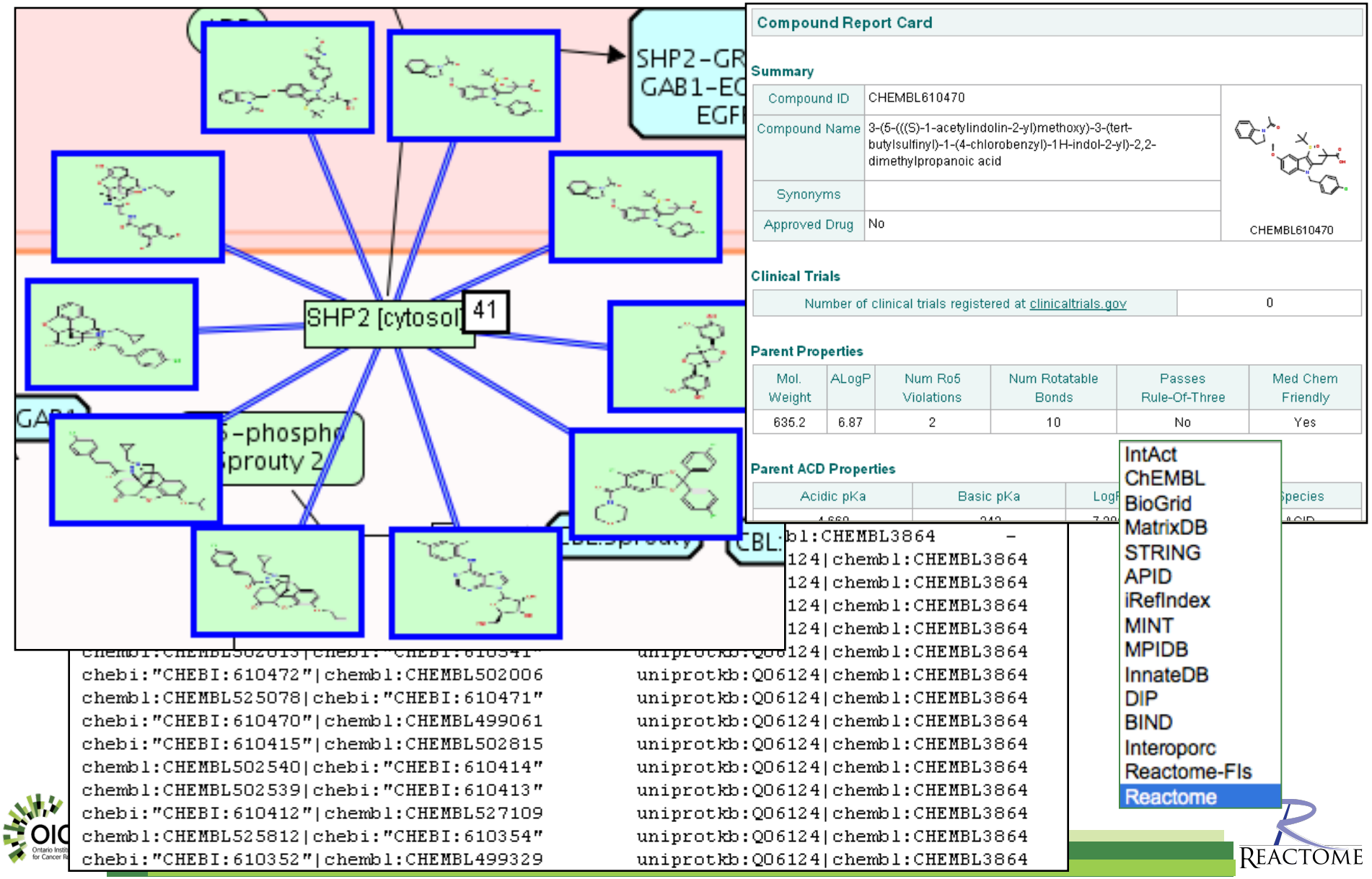
Signaling by constitutively active EGFR

Stable identifier	REACT_115852.2
Authored	Orlic-Milacic, M, 2011-11-04
Reviewed	Greulich, H, 2011-11-15 Savas, S, 2011-11-15

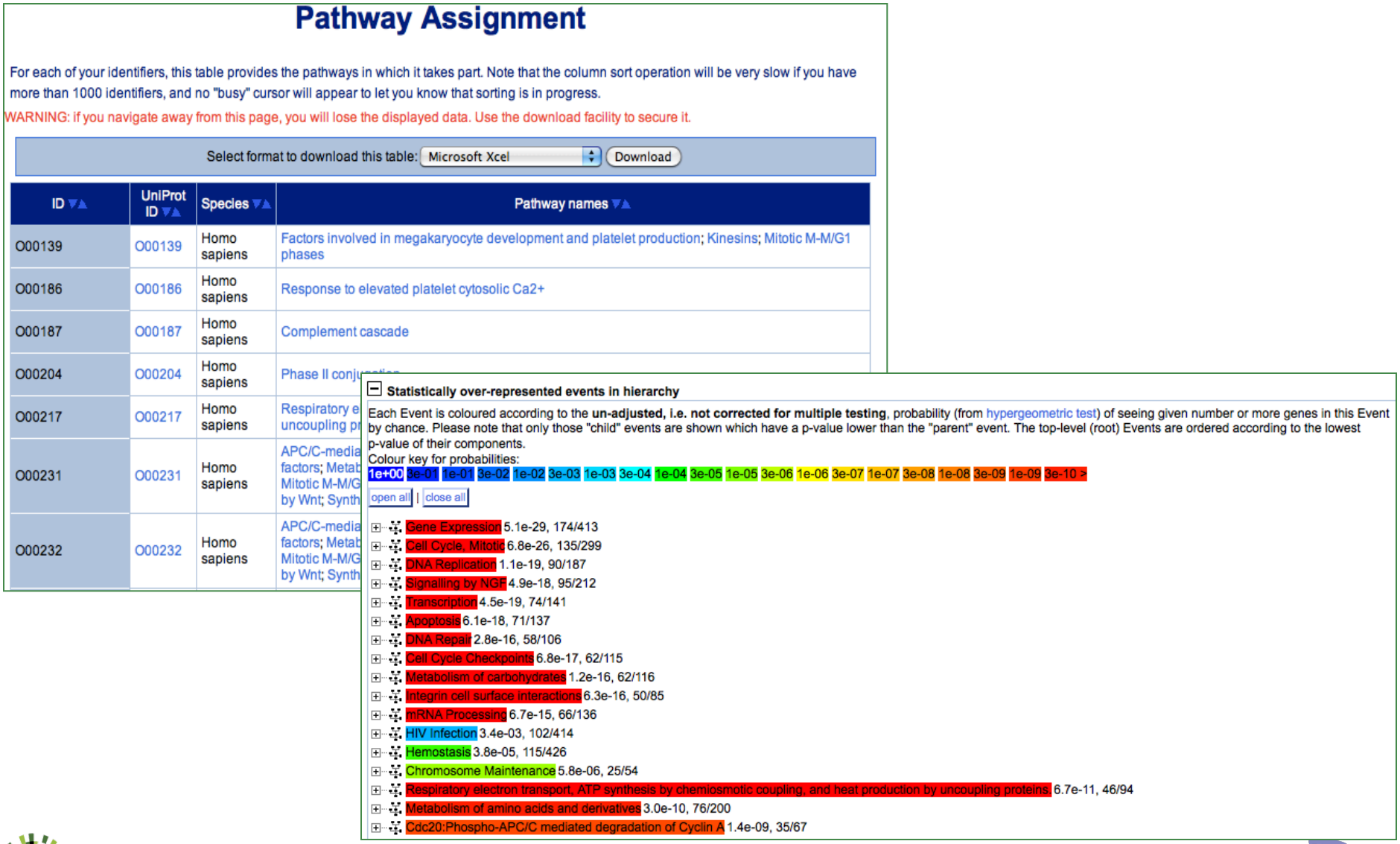
Signaling by EGFR is frequently activated in cancer through either genomic amplification of the EGFR locus, resulting in over-expression of the wild-type protein, or through activating mutations in the coding sequence of the EGFR gene, resulting in expression of a constitutively active mutant protein.

Epidermal growth factor receptor kinase domain mutants are present in ~16% of non-small-cell lung cancers (NSCLCs), but are also found in other cancer types, such as breast cancer, colorectal cancer, ovarian cancer and thyroid cancer. EGFR kinase domain mutants harbor activating mutations in exons 18-21 which code for the kinase domain (amino acids 712-979) . Small deletions, insertions or substitutions of amino acids within the kinase domain lock EGFR in its active conformation in which the enzyme can dimerize and undergo autophosphorylation spontaneously, without ligand binding (although ligand binding ability is preserved), and activate downstream signaling pathways that promote cell survival (Greulich et al. 2005, Zhang et al. 2006, Yun et al. 2007, Red Brewer et al. 2009).

Molecular Interaction Overlay – Small Molecule-protein Interactions



Pathway Analysis – ID Mapping & Overrepresentation



Pathway Assignment

For each of your identifiers, this table provides the pathways in which it takes part. Note that the column sort operation will be very slow if you have more than 1000 identifiers, and no "busy" cursor will appear to let you know that sorting is in progress.

WARNING: if you navigate away from this page, you will lose the displayed data. Use the download facility to secure it.

Select format to download this table: Microsoft Xcel Download

ID	UniProt ID	Species	Pathway names
O00139	O00139	Homo sapiens	Factors involved in megakaryocyte development and platelet production; Kinesins; Mitotic M-M/G1 phases
O00186	O00186	Homo sapiens	Response to elevated platelet cytosolic Ca2+
O00187	O00187	Homo sapiens	Complement cascade
O00204	O00204	Homo sapiens	Phase II conju
O00217	O00217	Homo sapiens	Respiratory e uncoupling pr
O00231	O00231	Homo sapiens	APC/C-media factors; Metab Mitotic M-M/G by Wnt; Synth
O00232	O00232	Homo sapiens	APC/C-media factors; Metab Mitotic M-M/G by Wnt; Synth

Statistically over-represented events in hierarchy

Each Event is coloured according to the **un-adjusted, i.e. not corrected for multiple testing**, probability (from hypergeometric test) of seeing given number or more genes in this Event by chance. Please note that only those "child" events are shown which have a p-value lower than the "parent" event. The top-level (root) Events are ordered according to the lowest p-value of their components.

Colour key for probabilities:

1e+00 3e-01 1e-01 3e-02 1e-02 3e-03 1e-03 3e-04 1e-04 3e-05 1e-05 3e-06 1e-06 3e-07 1e-07 3e-08 1e-08 3e-09 1e-09 3e-10 >

open all | close all

- Gene Expression 5.1e-29, 174/413
- Cell Cycle, Mitotic 6.8e-26, 135/299
- DNA Replication 1.1e-19, 90/187
- Signalling by NGF 4.9e-18, 95/212
- Transcription 4.5e-19, 74/141
- Apoptosis 6.1e-18, 71/137
- DNA Repair 2.8e-16, 58/106
- Cell Cycle Checkpoints 6.8e-17, 62/115
- Metabolism of carbohydrates 1.2e-16, 62/116
- Integrin cell surface interactions 6.3e-16, 50/85
- mRNA Processing 6.7e-15, 66/136
- HIV Infection 3.4e-03, 102/414
- Hemostasis 3.8e-05, 115/426
- Chromosome Maintenance 5.8e-06, 25/54
- Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins 6.7e-11, 46/94
- Metabolism of amino acids and derivatives 3.0e-10, 76/200
- Cdc20:Phospho-APC/C mediated degradation of Cyclin A 1.4e-09, 35/67

OICR Ontario Institute for Cancer Research

REACTOME

Species Comparison

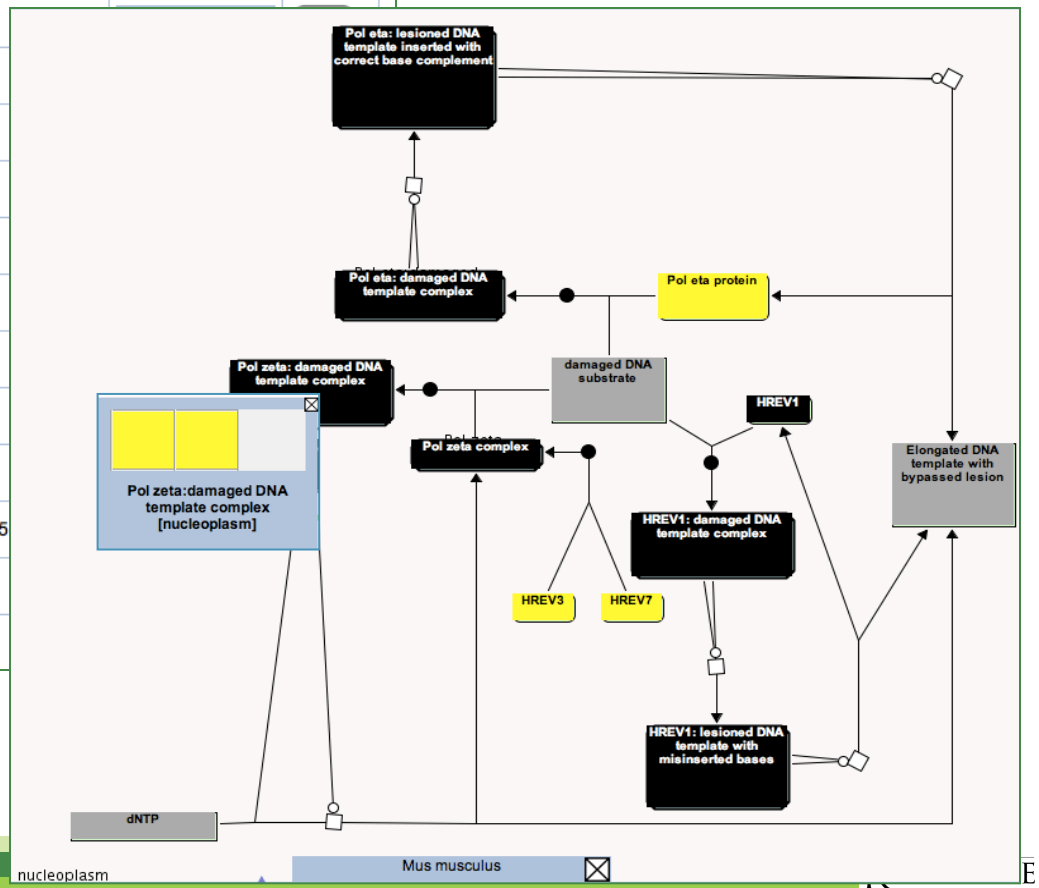
Species Comparison

This tool allows you to compare pathways between human and any of the other species inferred from Reactome by orthology. More....

WARNING: if you navigate away from this page, you will lose the displayed data. Use the download facility to secure it.

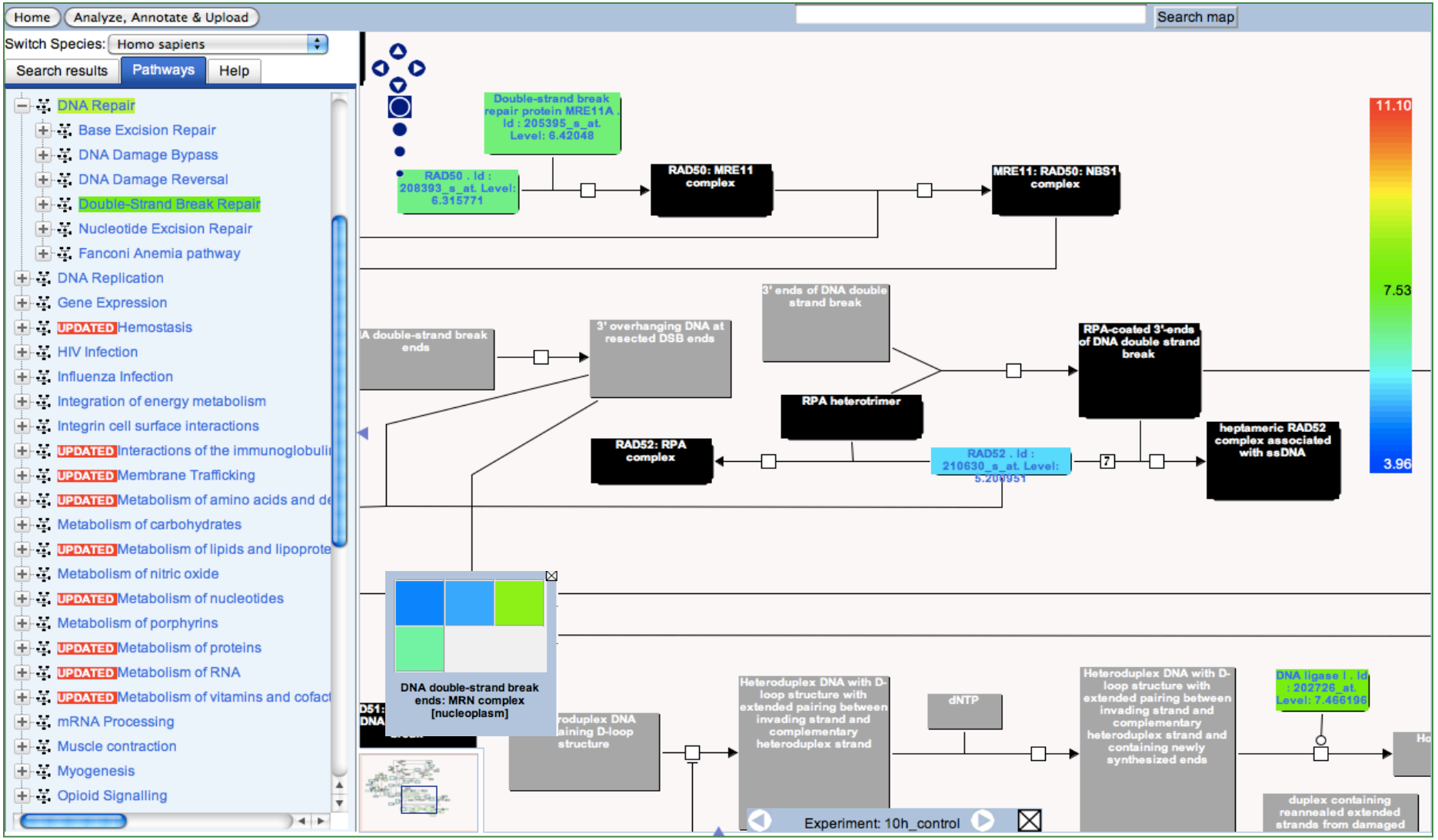
Select format to download this table: Microsoft Xcel Download

Pathway name	Other species	Proteins, human	Proteins, other species	% in other species	Click button to view pathway
ABC-family proteins mediated transport	Mus musculus	15	15	100%	View
Advanced glycosylation endproduct receptor signaling	Mus musculus	13	6		
APC/C-mediated degradation of cell cycle proteins	Mus musculus	79	74		
Apoptotic execution phase	Mus musculus	48	46		
Aquaporin-mediated transport	Mus musculus	30	21		
Asparagine N-linked glycosylation	Mus musculus	85	23		
Base Excision Repair	Mus musculus	19	19		
Bile acid and bile salt metabolism	Mus musculus	27	27		
Binding of RNA by Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs)	Mus musculus	3	3		
Botulinum neurotoxicity	Mus musculus	10	4		
Cell Cycle Checkpoints	Mus musculus	120	115		
Cell junction organization	Mus musculus	82	66		
Cell surface interactions at the vascular wall	Mus musculus	93	88		



OICR Ontario Institute for Cancer Research

Expression Analysis



Reactome data is open access and open source

The whole content of the Reactome can be downloaded as:

- MySQL database dump (To use skypainter you also need this database)

MySQL db

Several narrower datasets and download formats are also available:

- Human protein-protein interaction pairs in tab-delimited format.
- Full list of protein-protein interactions, including non-human species.
- Human protein-protein interaction pairs in PSI-MITAB format.

Interactions

- Uniprot to pathways plus stable id mappings
- Uniprot to pathways mappings
- Curated and inferred Uniprot to pathways mappings
- Gene association file (GO annotation in Reactome)

Mapping

- Human reactions in SBML (level 2, version 3) format.

SBML

- Events in the **BioPAX level 2** format
- Events in the **BioPAX level 3** format

BioPAX

- Reactome Pathways Gene Set
- Protégé 2.0 ontology for Reactome data model: **pins file**, **pont file**, **pprj file**
- Human pathway diagrams: PDF, or PNG

GSEA

The complete Reactome textbook of biological pathways and processes can be downloaded in:

- PDF format (ZIP archive).
- RTF format (ZIP archive, split into files on a per-chapter basis).

Textbook

The Reactome website can be installed locally:

- Instructions for local installation of Reactome database and website.
- Download Reactome website contents and perl code (requires additional n

Local install

Two **Java tools** are available for Reactome data entry. The **Author tool** is designed for biologists to input data into Reactome, while the **Curator tool** is meant for use by curators to annotate biological pathways based on the Reactome schema. N.B. these tools will only wor with a database using our internal database format, the database available from this web page will **not** work with them. Click on the help link at the top of this page to request a copy of our internal database.

- Curator Tool (version 3.1, build 66, January 24, 2012)

Curator tool

- Author Tool Version 4.0 in Java Web Start
- Author Tool Version 4.0 in Java Applet
- Author Tool (version 2.1, build 25, for some old projects)

Author tool

A list of **functional interactionos (FIs)** can be downloaded in three formats. For details about FIs, please see Wu G, Feng X, Stein L.

- MySQL database dump
- Tab delimited text files
- PSI-MITAB format

FIs

A **SOAP** based Web Services API is available to access the Reactome data. For details about this API, please follow the following links:

- Simple Description for the Reactome Web Services API
- Training Materials for the Reactome Web Services API
  - Reactome SOAP WS User's Guide in PDF (1M)
  - Reactome SOAP WS Tutorial in Power Point Slides (2M)
  - Reactome SOAP WS Tutorial in Flash Movie (640 x 480) (11M)
  - Reactome SOAP WS Tutorial in Flash Movie (800 x 600) (12M)
- XML Schema for the data model
- WSDL file for the Reactome Web Servies API

SOAP/APIs

A **RESTful** API (**beta**) is available too to access the Reactome data. For details about this API, please see this document: Reactome RESTful API

OICR Ontario Institute for Cancer Research

REACTOME

Programmatic Access

- Application programming interfaces (API) are important to connect and automate data exchange between local programs and databases.
  - BioMart API
  - MySQL/Perl API
  - MySQL/Java API
  - SOAP/WSDL
  - RESTful API (beta)

Continuing Priorities…….

- Expand curated pathway content
  - Model organism
  - Disease, e.g. cancer, cardiovascular, infectious, etc.
  - Work with the community for improved ontology/link out support
    - GO, PRO, MGI, RGD, DO, etc.

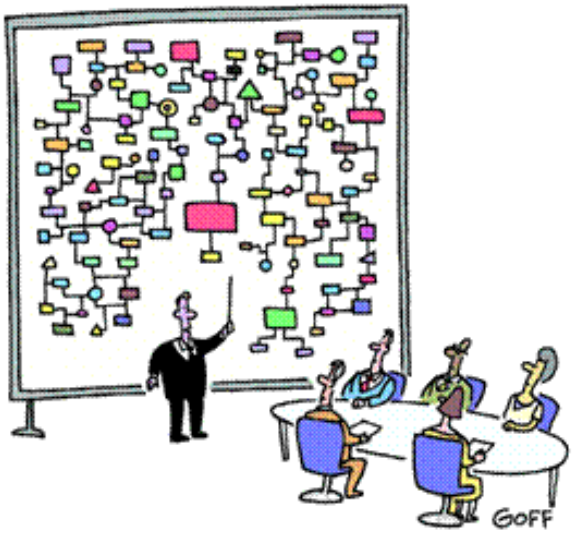
- Develop new visualization and analysis tools
  - HMTL5 canvas based pathway browser
  - Bioconductor and Galaxy pipelines for Pathway analysis
  - Pathway tracks for Genome Browsers
  - Mobile-friendly Reactome
  - Reactome in the Cloud

Summary

- Pathway databases are an integral part of the scientific community.
- Reactome is a highly reliable, human-curated database of biological pathways and reactions.
- Deployed a user-friendly web site and tools that promotes integrated research on pathways and networks.
  - Data visualization
  - Data analysis
  - Data expansion
  - Data integration
  - Data standards
  - Data exports
- Develop and distribute open software and standard operating procedures for the management of pathway information.

Thanks

- Marija Orlic-Milacic
- Karen Rothfels
- Peter D'Eustachio
- Lisa Matthews
- Marc Gillespie
- Guanming Wu
- Lincoln Stein
- Michael Caudy
- David Croft
- Phani Garapati
- Henning Hermjakob
- Bijay Jassal
- Steve Jupe
- Bruce May
- Antonio Fabregat Mundo
- Veronica Shamovsky
- Joel Weiser
- Mark Williams



"And that's why we need a computer."

CSH EMBL-EBI National Human Genome Research Institute

NYU School of Medicine Ontario Ministry of Economic Development and Innovation

OICR Ontario Institute for Cancer Research

REACTOME