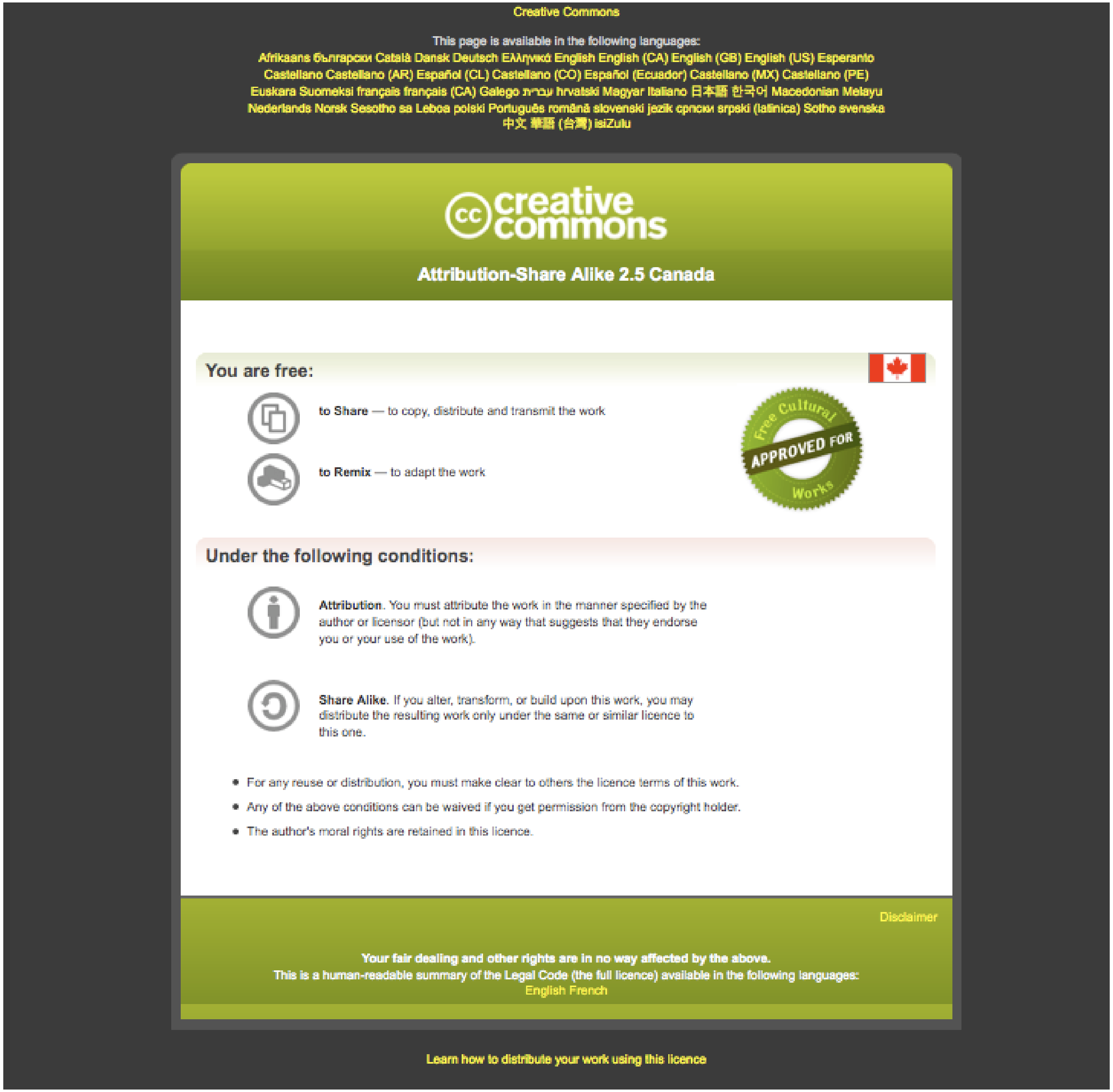
廣州醫科大學
GUANGZHOU MEDICAL UNIVERSITY

Gene Set Analysis –Methods and Tools

Antonio Mora, Ph.D.
(Wechat: antoniocmora)

20.12.2018

MORA LAB
FUNCTIONAL BIOINFORMATICS



Creative Commons

This page is available in the following languages:
Afrikaans български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE) Euskara Suomeksi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska 中文 華語 (台灣) isiZulu

creative commons

Attribution-Share Alike 2.5 Canada

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 or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

Disclaimer

Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
English French

Learn how to distribute your work using this licence

Contents

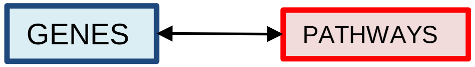
1.1. Introduction. Biological annotation sources.
1.2. Pathway databases
1.3. Pathway visualization
1.4. Gene Ontology
1.5. Gene Set databases
1.6. Automatic reconstruction of pathways

1.1. Introduction

Databases are sources of Biological Annotation

Pathway databases

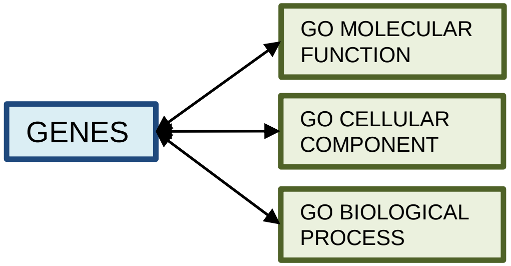
(KEGG, Reactome, Wikipathways)



GENES
PATHWAYS

Ontology databases

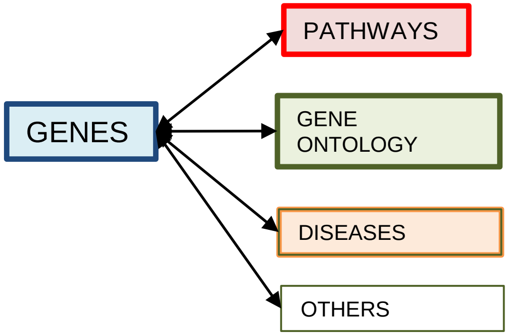
(Gene Ontology)



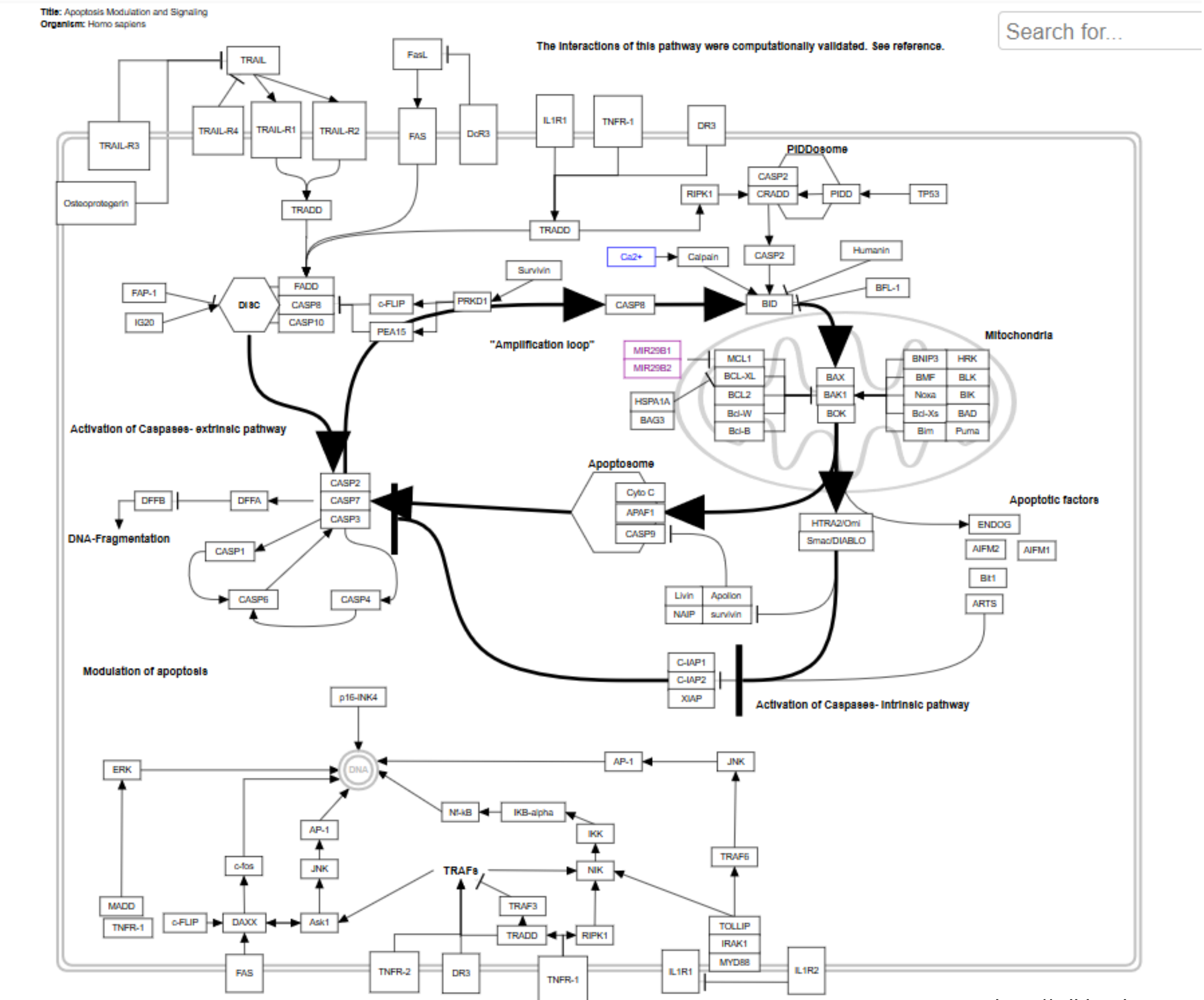
GENES
GO MOLECULAR FUNCTION
GO CELLULAR COMPONENT
GO BIOLOGICAL PROCESS

Gene set databases

(GeneSetDB, MSigDB)



GENES
PATHWAYS
GENE ONTOLOGY
DISEASES
OTHERS

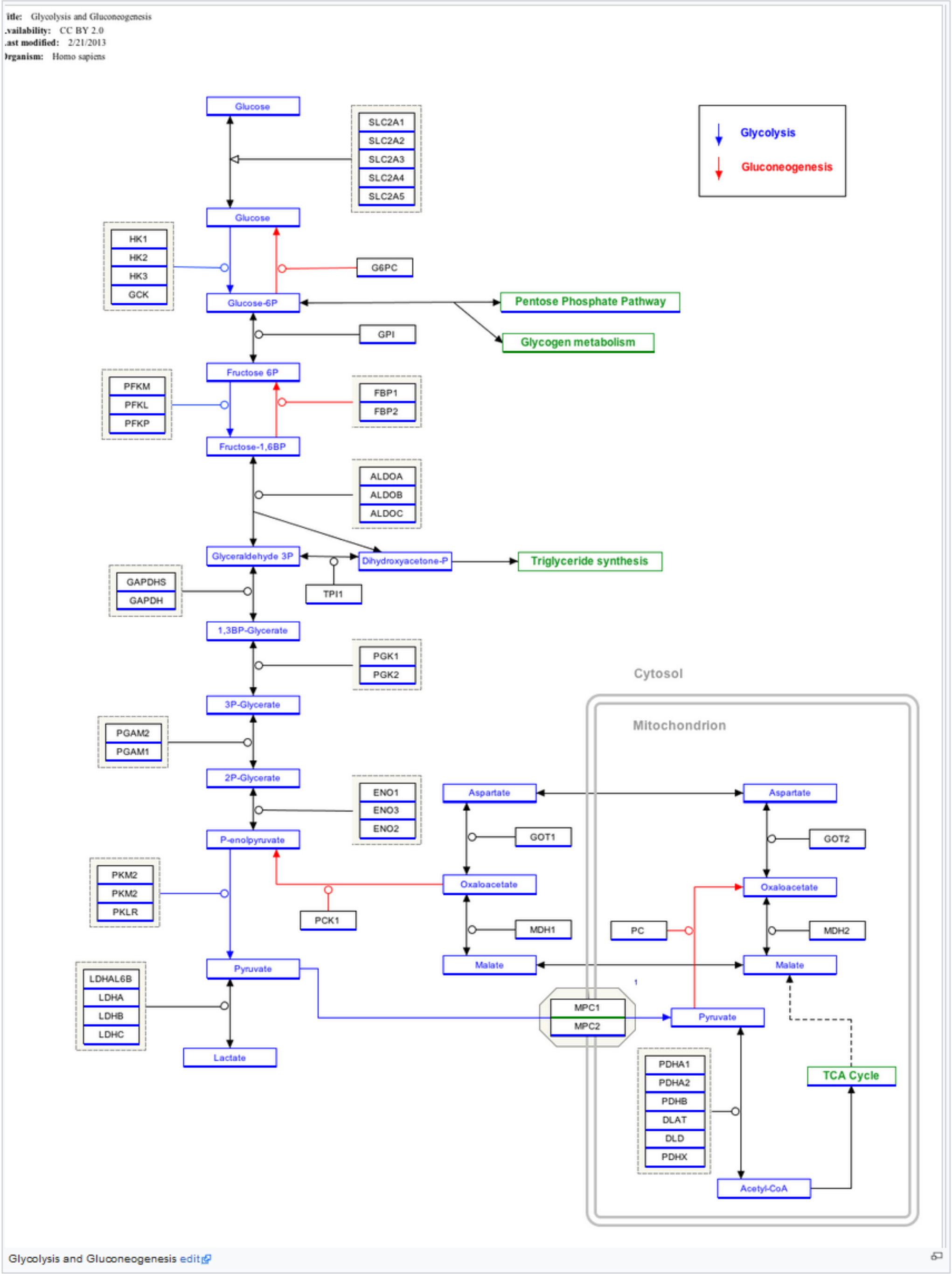


http://wikipathways.org/index.php/Pathway:WP1772

A ***biological pathway*** is a series of interactions or chemical reactions among molecules that leads to one or more products.

The most studied types of biological pathways are: Metabolic pathways, signal transduction pathways, and gene regulation pathways.

Metabolic pathways: Glycolysis

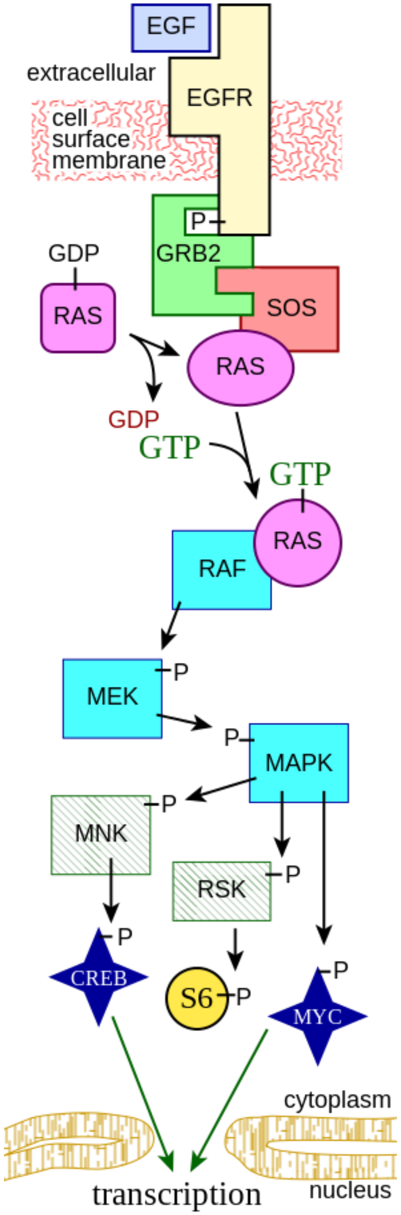


itle: Glycolysis and Gluconeogenesis
.vailability: CC BY 2.0
.ast modified: 2/21/2013
)rganism: Homo sapiens

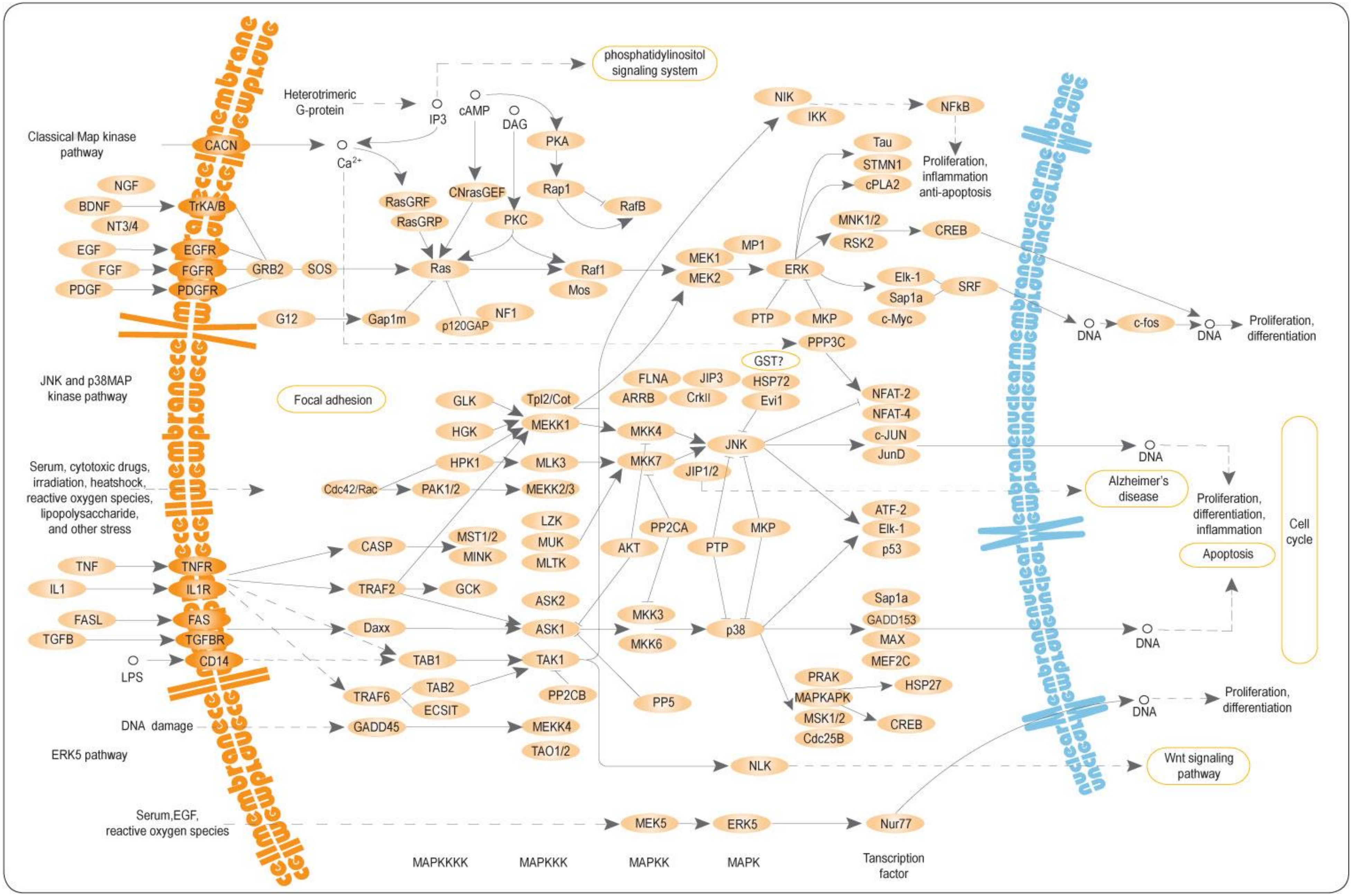
Glycolysis and Gluconeogenesis edit

https://en.wikipedia.org/wiki/Glycolysis

Signaling pathways: MAPK/ERK pathway



https://en.wikipedia.org/wiki/MAPK/ERK_pathway



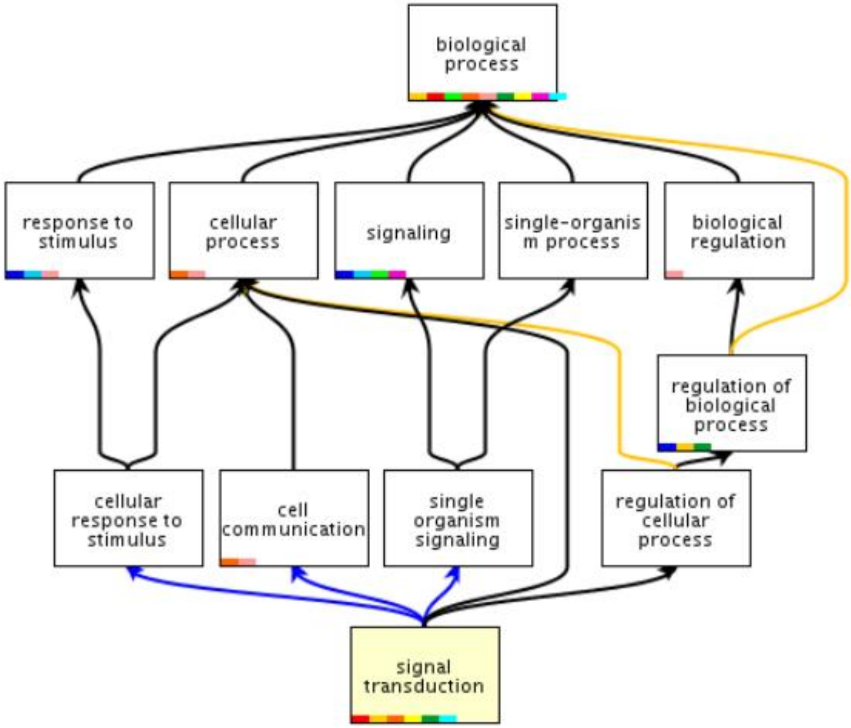
https://en.wikipedia.org/wiki/MAPK/ERK_pathway

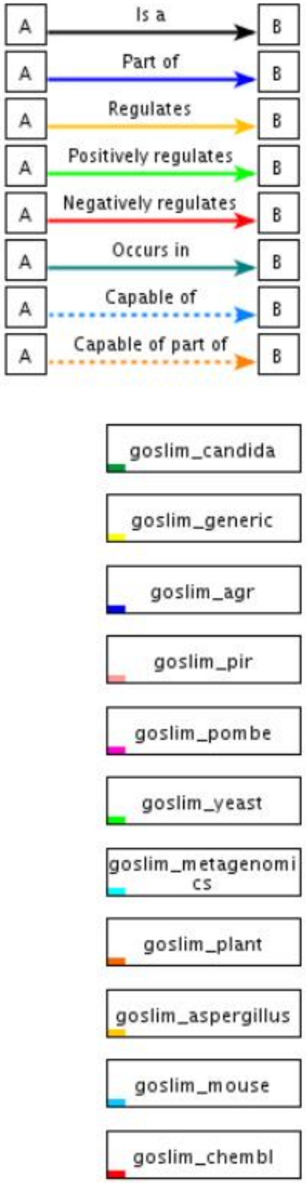
A precise definition of a pathway?: The definition of a pathway is a little subjective. Three problems:

- Where to start and where to end
- Level of detail (intermediate reactions)
- Pathway cross-talk

Therefore, pathways may look slightly different according to the source.

Ontologies

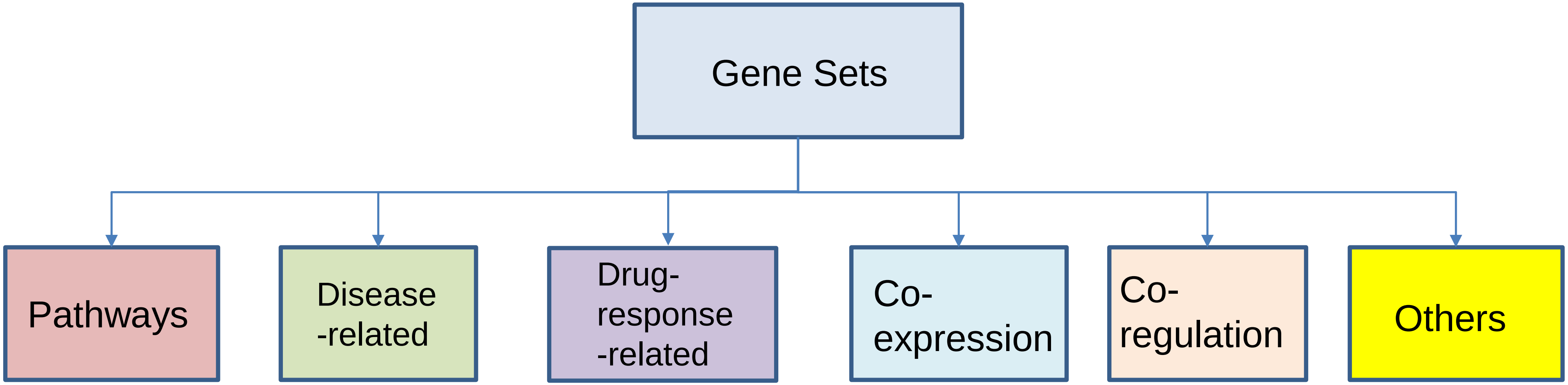




An ***ontology*** is a way of organizing the knowledge in a field. Knowledge is organized in terms of all of the concepts involved, and a graph of the way in which such concepts relate to each other.

QuickGO - http://www.ebi.ac.uk/QuickGO

Gene sets



Essentially, any set of genes that can be grouped for some reason.

1.2. Pathway databases.

KEGG

Reactome

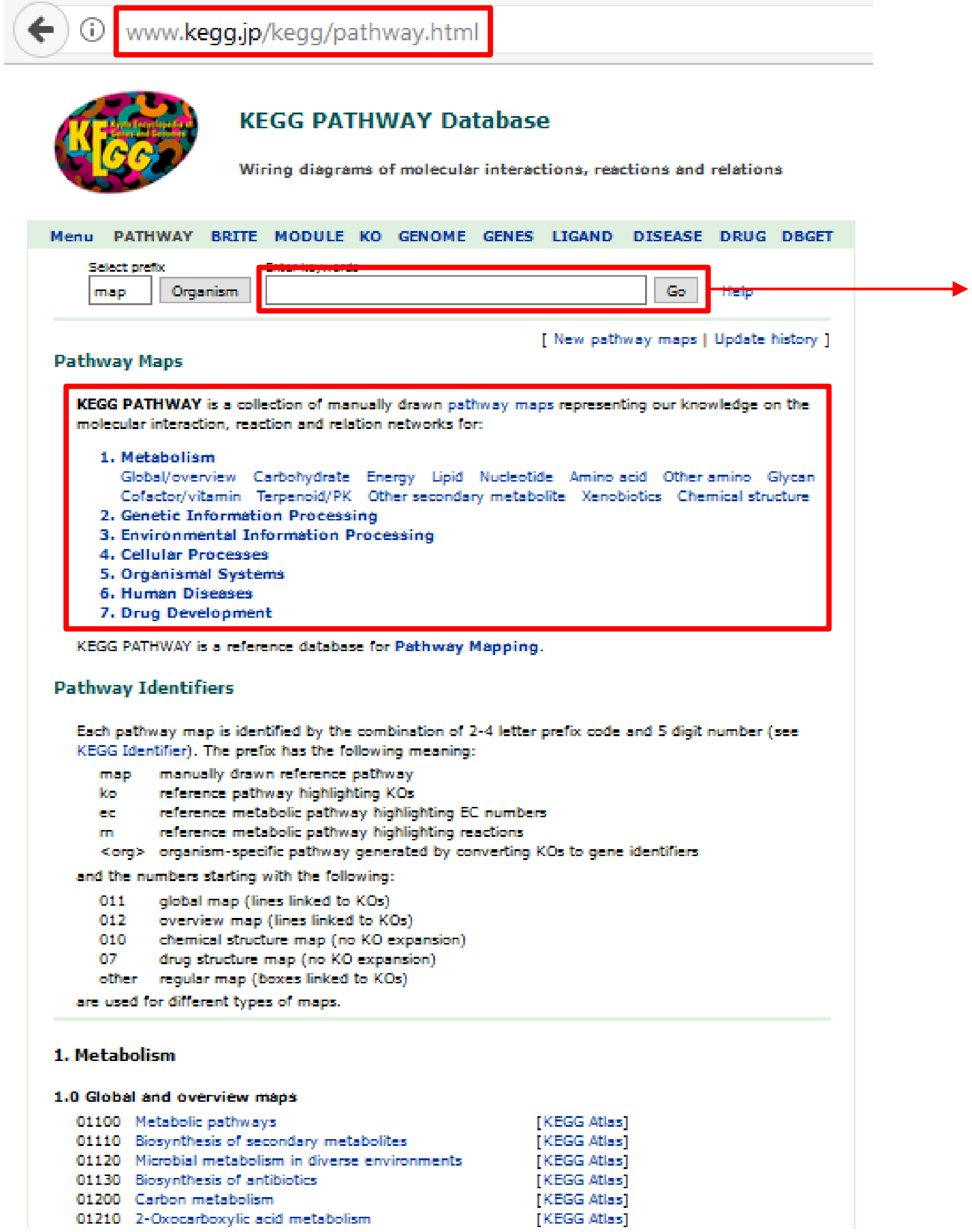
Wikipathways

Pathway Commons

Biocyc

Panther Pathway

Pathway databases: KEGG



www.kegg.jp/kegg/pathway.html

KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions and relations

Menu PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET
Select prefix
map
Organism
Enter keywords
Go
Help

Search

[New pathway maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction, reaction and relation networks for:

1. Metabolism
   Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan
   Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Organismal Systems
6. Human Diseases
7. Drug Development

KEGG PATHWAY is a reference database for Pathway Mapping.

Pathway Identifiers

Each pathway map is identified by the combination of 2-4 letter prefix code and 5 digit number (see KEGG Identifier). The prefix has the following meaning:

map	manually drawn reference pathway
ko	reference pathway highlighting KOs
ec	reference metabolic pathway highlighting EC numbers
rn	reference metabolic pathway highlighting reactions
<org>	organism-specific pathway generated by converting KOs to gene identifiers

and the numbers starting with the following:

011	global map (lines linked to KOs)
012	overview map (lines linked to KOs)
010	chemical structure map (no KO expansion)
07	drug structure map (no KO expansion)
other	regular map (boxes linked to KOs)

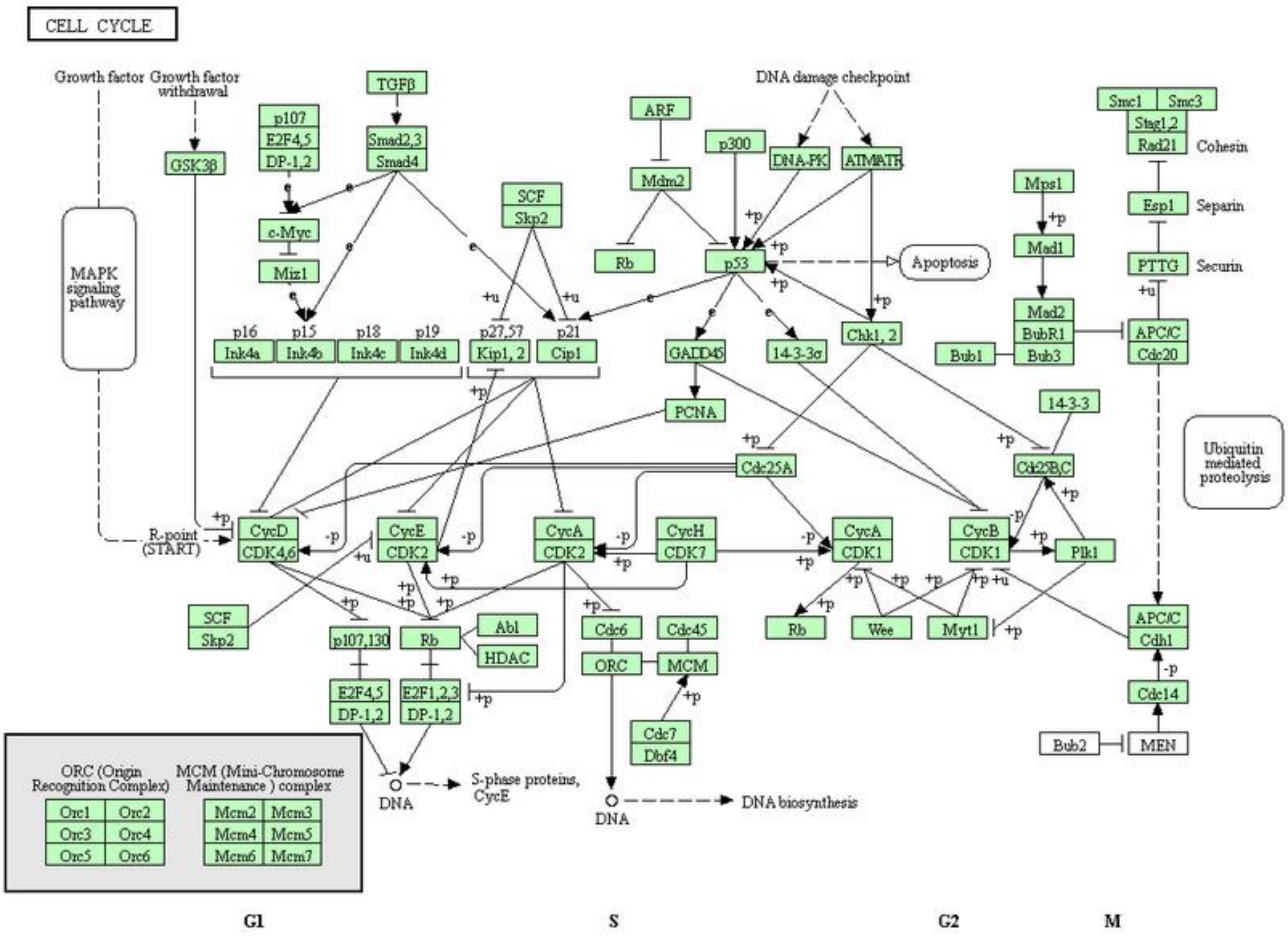
are used for different types of maps.

1. Metabolism

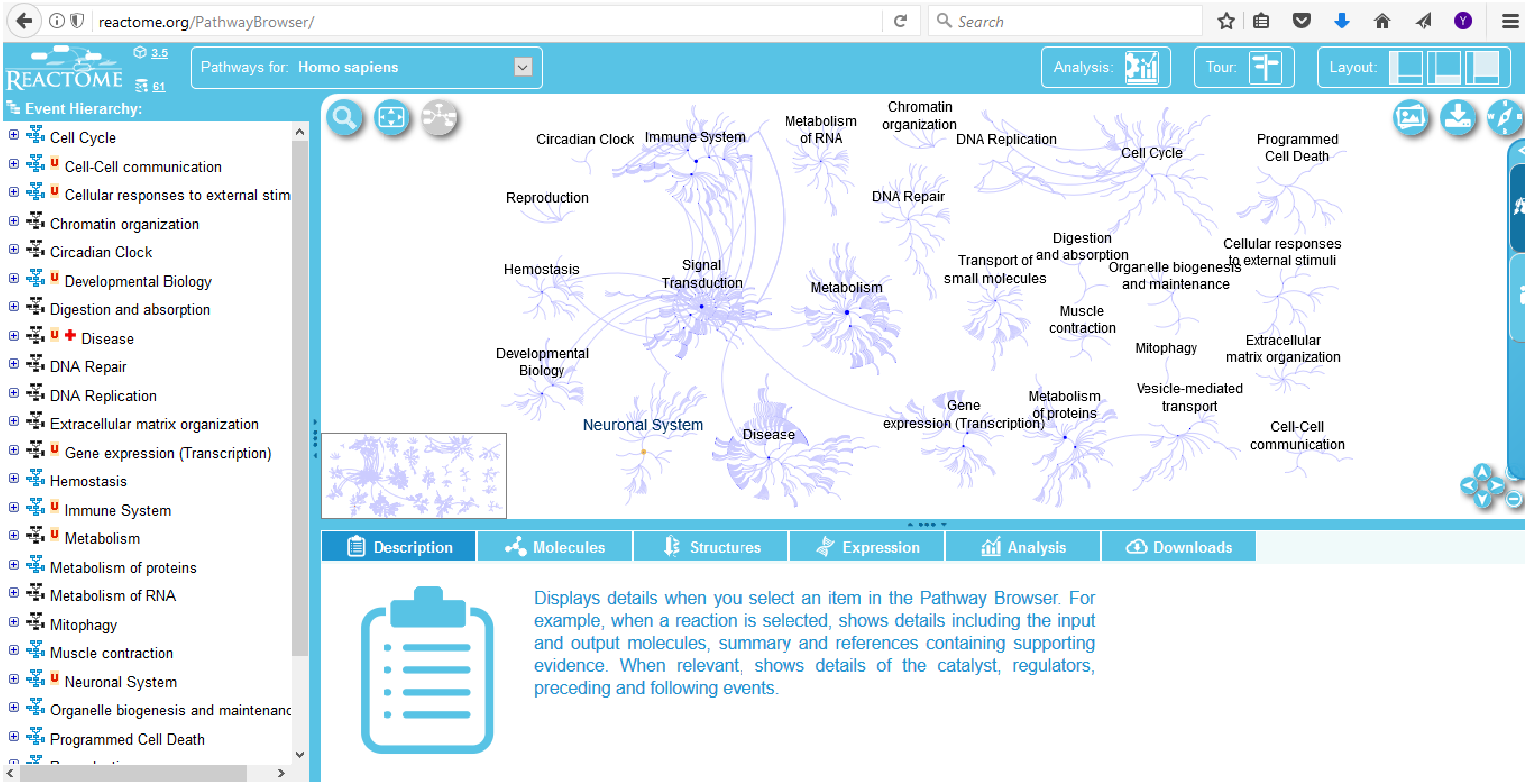
1.0 Global and overview maps

01100	Metabolic pathways	[KEGG Atlas]
01110	Biosynthesis of secondary metabolites	[KEGG Atlas]
01120	Microbial metabolism in diverse environments	[KEGG Atlas]
01130	Biosynthesis of antibiotics	[KEGG Atlas]
01200	Carbon metabolism	[KEGG Atlas]
01210	2-Oxocarboxylic acid metabolism	[KEGG Atlas]

Pathway databases: KEGG



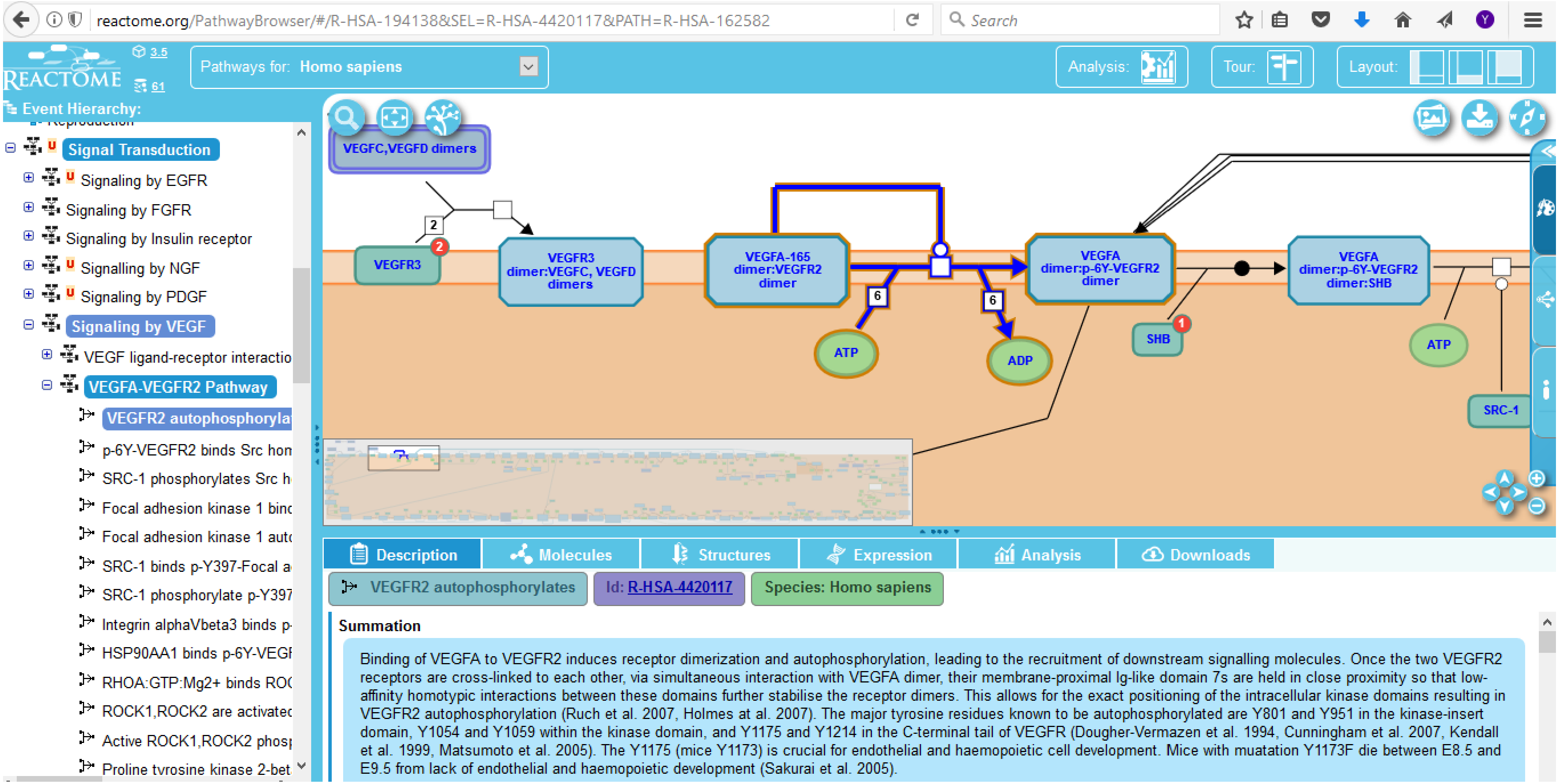
Pathway databases: Reactome



- Cell Cycle
- Cell-Cell communication
- Cellular responses to external stim
- Chromatin organization
- Circadian Clock
- Developmental Biology
- Digestion and absorption
- Disease
- DNA Repair
- DNA Replication
- Extracellular matrix organization
- Gene expression (Transcription)
- Hemostasis
- Immune System
- Metabolism
- Metabolism of proteins
- Metabolism of RNA
- Mitophagy
- Muscle contraction
- Neuronal System
- Organelle biogenesis and maintenanc
- Programmed Cell Death

Displays details when you select an item in the Pathway Browser. For example, when a reaction is selected, shows details including the input and output molecules, summary and references containing supporting evidence. When relevant, shows details of the catalyst, regulators, preceding and following events.

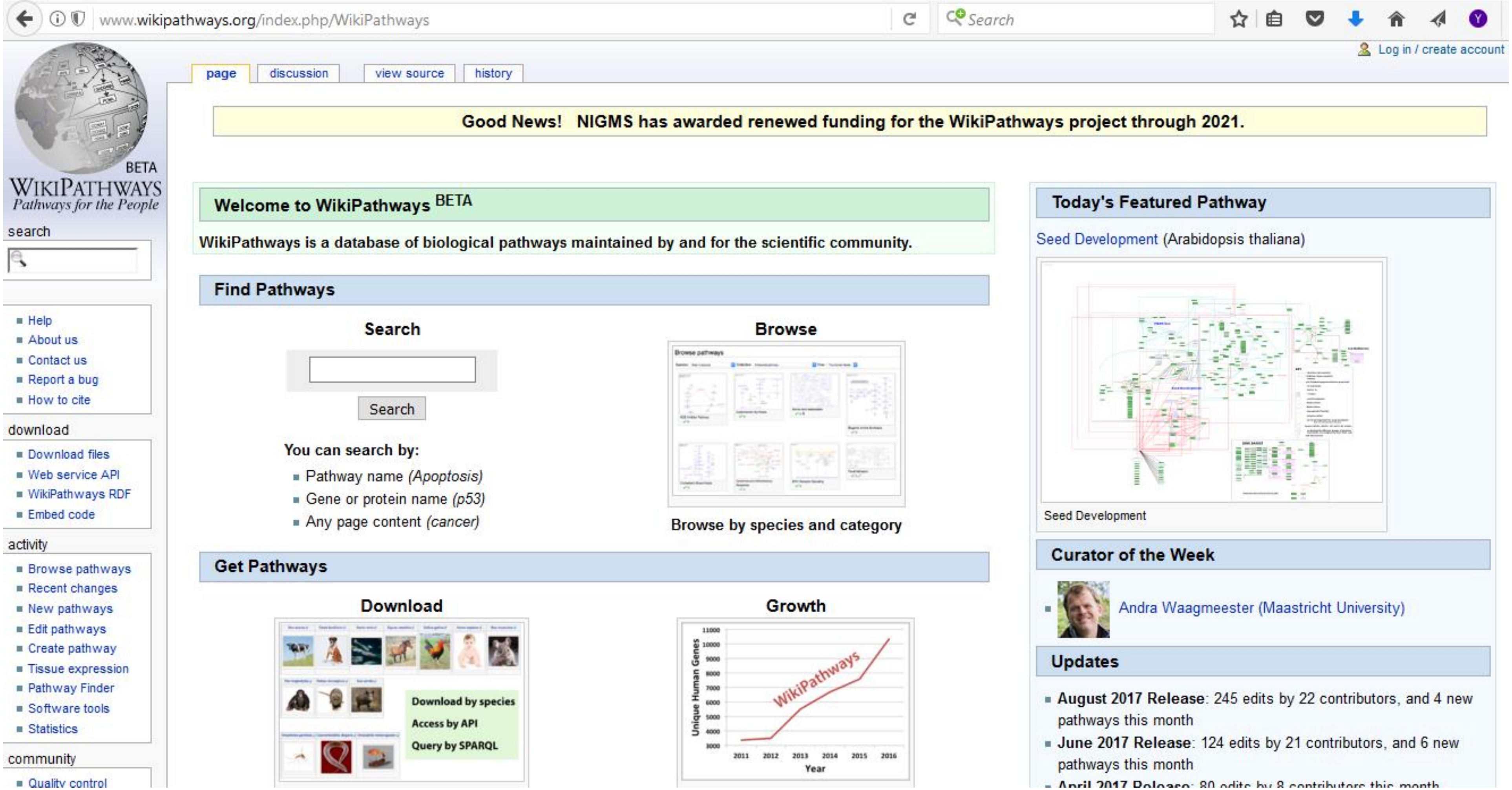
Pathway databases: Reactome


reactome.org/PathwayBrowser/#/R-HSA-194138&SEL=R-HSA-4420117&PATH=R-HSA-162582
Search
REACTOME
3.5
61
Pathways for: Homo sapiens
Analysis:
Tour:
Layout:
Event Hierarchy:
Reproduction
Signal Transduction
Signaling by EGFR
Signaling by FGFR
Signaling by Insulin receptor
Signalling by NGF
Signaling by PDGF
Signaling by VEGF
VEGF ligand-receptor interactio
VEGFA-VEGFR2 Pathway
VEGFR2 autophosphoryla
p-6Y-VEGFR2 binds Src hon
SRC-1 phosphorylates Src h
Focal adhesion kinase 1 binc
Focal adhesion kinase 1 autc
SRC-1 binds p-Y397-Focal a
SRC-1 phosphorylate p-Y397
Integrin alphaVbeta3 binds p-
HSP90AA1 binds p-6Y-VEGF
RHOA:GTP:Mg2+ binds ROC
ROCK1,ROCK2 are activatec
Active ROCK1,ROCK2 phos
Proline tvrosine kinase 2-bet
VEGFC,VEGFD dimers
VEGFR3
2
VEGFR3 dimer:VEGFC, VEGFD dimers
VEGFA-165 dimer:VEGFR2 dimer
6
ATP
ADP
VEGFA dimer:p-6Y-VEGFR2 dimer
SHB
1
VEGFA dimer:p-6Y-VEGFR2 dimer:SHB
ATP
SRC-1
Description
Molecules
Structures
Expression
Analysis
Downloads
VEGFR2 autophosphorylates
Id: R-HSA-4420117
Species: Homo sapiens

Summation

Binding of VEGFA to VEGFR2 induces receptor dimerization and autophosphorylation, leading to the recruitment of downstream signalling molecules. Once the two VEGFR2 receptors are cross-linked to each other, via simultaneous interaction with VEGFA dimer, their membrane-proximal Ig-like domain 7s are held in close proximity so that low-affinity homotypic interactions between these domains further stabilise the receptor dimers. This allows for the exact positioning of the intracellular kinase domains resulting in VEGFR2 autophosphorylation (Ruch et al. 2007, Holmes at al. 2007). The major tyrosine residues known to be autophosphorylated are Y801 and Y951 in the kinase-insert domain, Y1054 and Y1059 within the kinase domain, and Y1175 and Y1214 in the C-terminal tail of VEGFR (Dougher-Vermazen et al. 1994, Cunningham et al. 2007, Kendall et al. 1999, Matsumoto et al. 2005). The Y1175 (mice Y1173) is crucial for endothelial and haemopoietic cell development. Mice with muatation Y1173F die between E8.5 and E9.5 from lack of endothelial and haemopoietic development (Sakurai et al. 2005).

Pathway databases: Wikipathways



www.wikipathways.org/index.php/WikiPathways

Log in / create account

page | discussion | view source | history

BETA
WIKIPATHWAYS
Pathways for the People

search

- Help
- About us
- Contact us
- Report a bug
- How to cite

download

- Download files
- Web service API
- WikiPathways RDF
- Embed code

activity

- Browse pathways
- Recent changes
- New pathways
- Edit pathways
- Create pathway
- Tissue expression
- Pathway Finder
- Software tools
- Statistics

community

- Quality control

Good News! NIGMS has awarded renewed funding for the WikiPathways project through 2021.

Welcome to WikiPathways BETA

WikiPathways is a database of biological pathways maintained by and for the scientific community.

Find Pathways

Search

Search

You can search by:

- Pathway name *(Apoptosis)*
- Gene or protein name *(p53)*
- Any page content *(cancer)*

Browse

Browse by species and category

Get Pathways

Download

Download by species
Access by API
Query by SPARQL

Growth

Today's Featured Pathway

Seed Development (Arabidopsis thaliana)

Seed Development

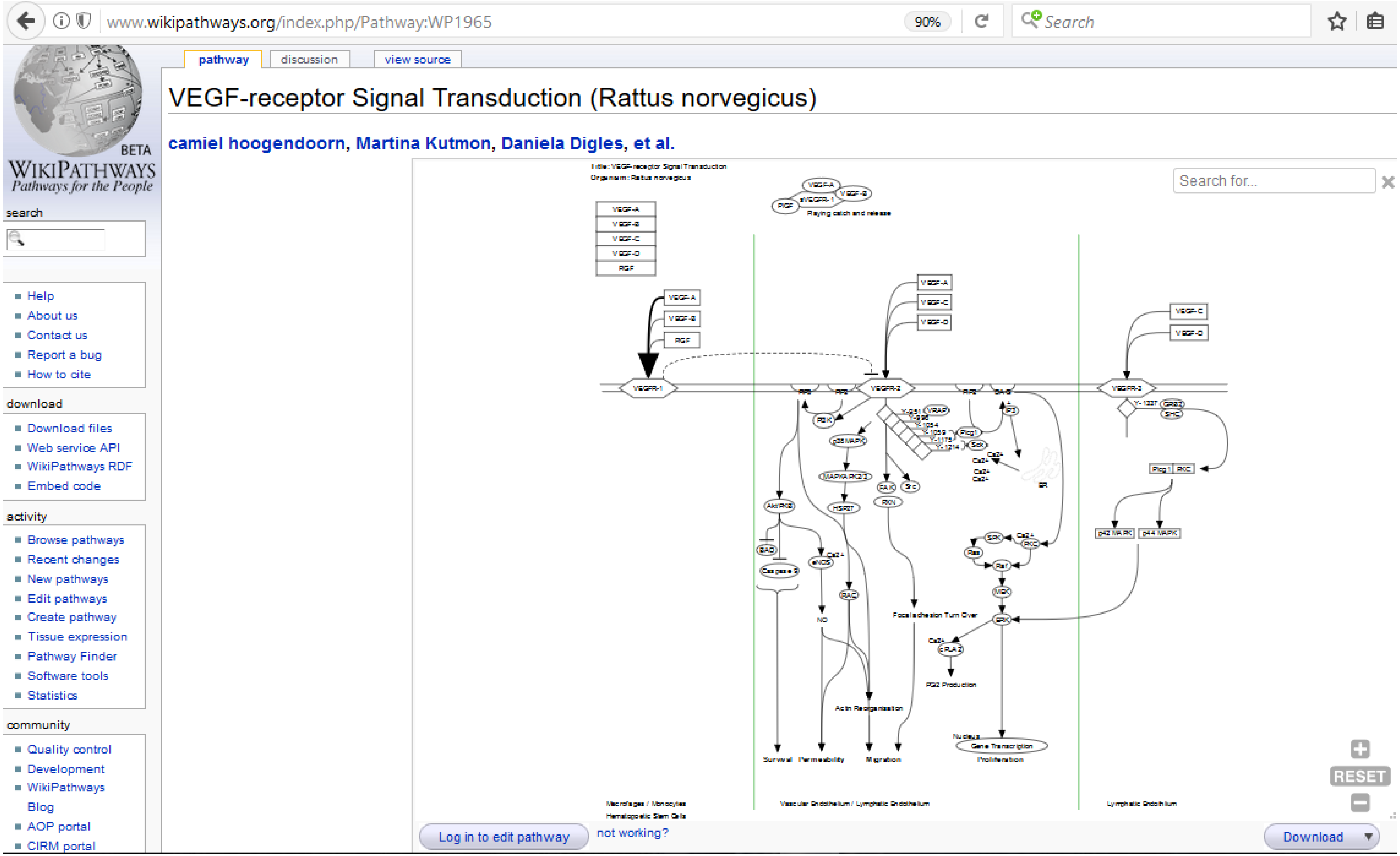
Curator of the Week

- Andra Waagmeester (Maastricht University)

Updates

- **August 2017 Release**: 245 edits by 22 contributors, and 4 new pathways this month
- **June 2017 Release**: 124 edits by 21 contributors, and 6 new pathways this month
- **April 2017 Release**: 80 edits by 8 contributors this month

Pathway databases: Wikipathways



www.wikipathways.org/index.php/Pathway:WP1965

VEGF-receptor Signal Transduction (Rattus norvegicus)

camiel hoogendoorn, Martina Kutmon, Daniela Digles, et al.

- Help
- About us
- Contact us
- Report a bug
- How to cite

download

- Download files
- Web service API
- WikiPathways RDF
- Embed code

activity

- Browse pathways
- Recent changes
- New pathways
- Edit pathways
- Create pathway
- Tissue expression
- Pathway Finder
- Software tools
- Statistics

community

- Quality control
- Development
- WikiPathways Blog
- AOP portal
- CIRM portal

Pathway databases: Reactome vs Wikipathways

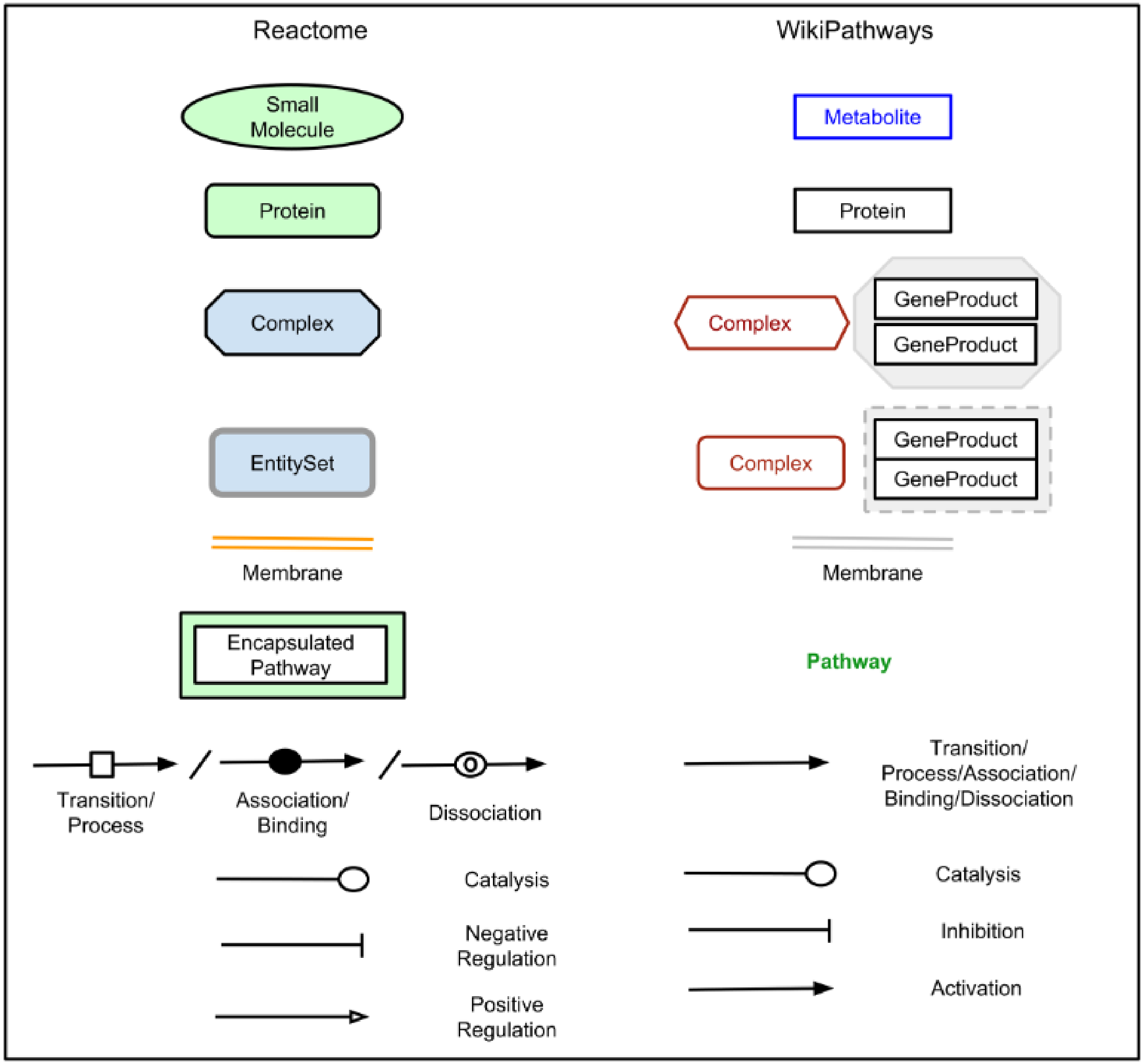
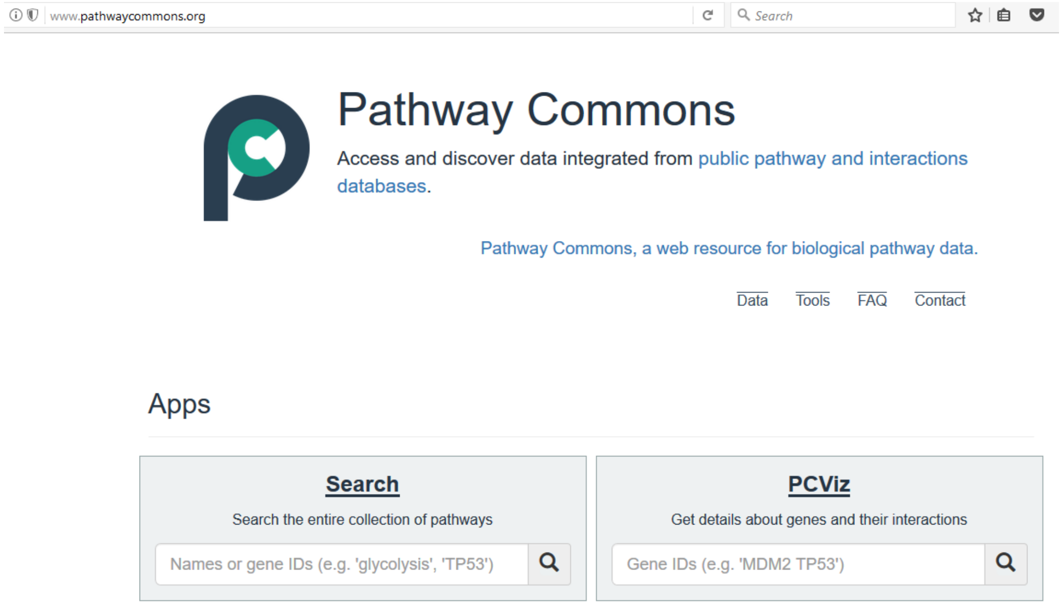


Fig 1. Mapping Reactome pathways elements to WikiPathways pathway elements. This diagram shows the symbols used to represent different biological entities in Reactome and the corresponding symbol used to represent the same biological entity in WikiPathways.

doi:10.1371/journal.pcbi.1004941.g001

Pathway databases: Pathway Commons

Search for pathways in multiple pathway databases



www.pathwaycommons.org

Pathway Commons

Access and discover data integrated from public pathway and interactions databases.

Pathway Commons, a web resource for biological pathway data.

Data Tools FAQ Contact

Apps

Search

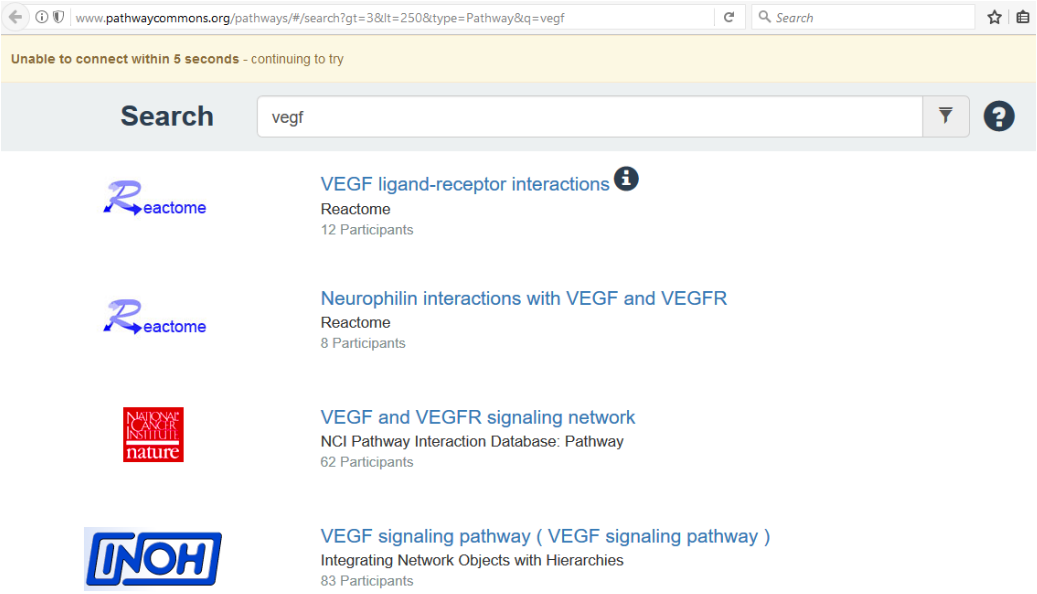
Search the entire collection of pathways

Names or gene IDs (e.g. 'glycolysis', 'TP53')

PCViz

Get details about genes and their interactions

Gene IDs (e.g. 'MDM2 TP53')



www.pathwaycommons.org/pathways/#/search?gt=3&lt=250&type=Pathway&q=vegf

Unable to connect within 5 seconds - continuing to try

Search vegf

VEGF ligand-receptor interactions
Reactome
12 Participants

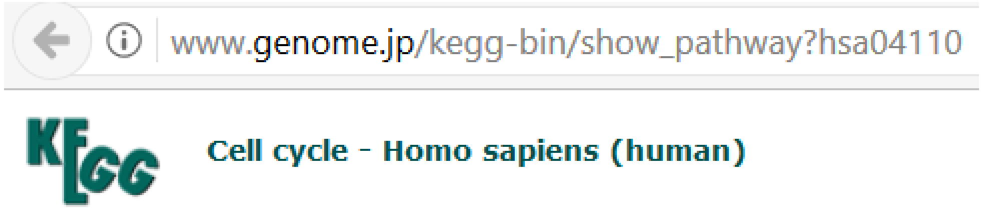
Neurophilin interactions with VEGF and VEGFR
Reactome
8 Participants

VEGF and VEGFR signaling network
NCI Pathway Interaction Database: Pathway
62 Participants

VEGF signaling pathway (VEGF signaling pathway)
Integrating Network Objects with Hierarchies
83 Participants

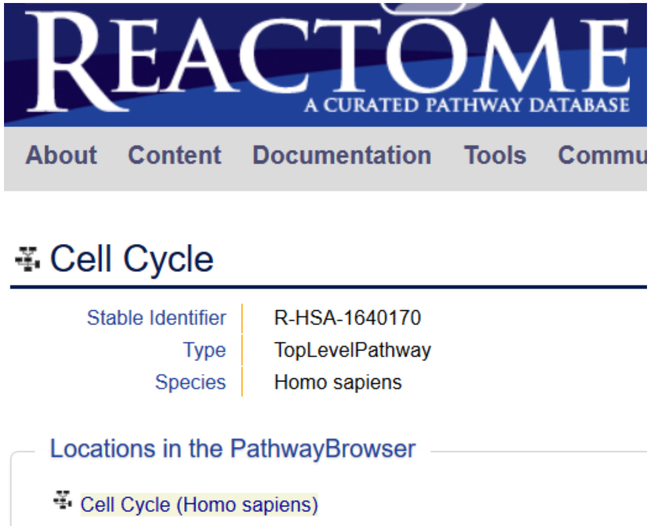
Be aware of... Pathway database identifiers

Identifiers (IDs) are ideally unique, stable names or numbers that help track database records. For example, your wechat ID, Entrez Gene ID 41232, etc. Each DB has its own type of identifier.



www.genome.jp/kegg-bin/show_pathway?hsa04110

Cell cycle - Homo sapiens (human)

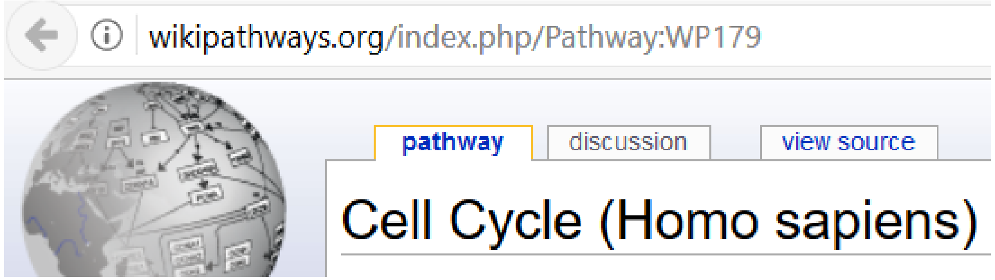


Cell Cycle

Stable Identifier	R-HSA-1640170
Type	TopLevelPathway
Species	Homo sapiens

Locations in the PathwayBrowser

Cell Cycle (Homo sapiens)



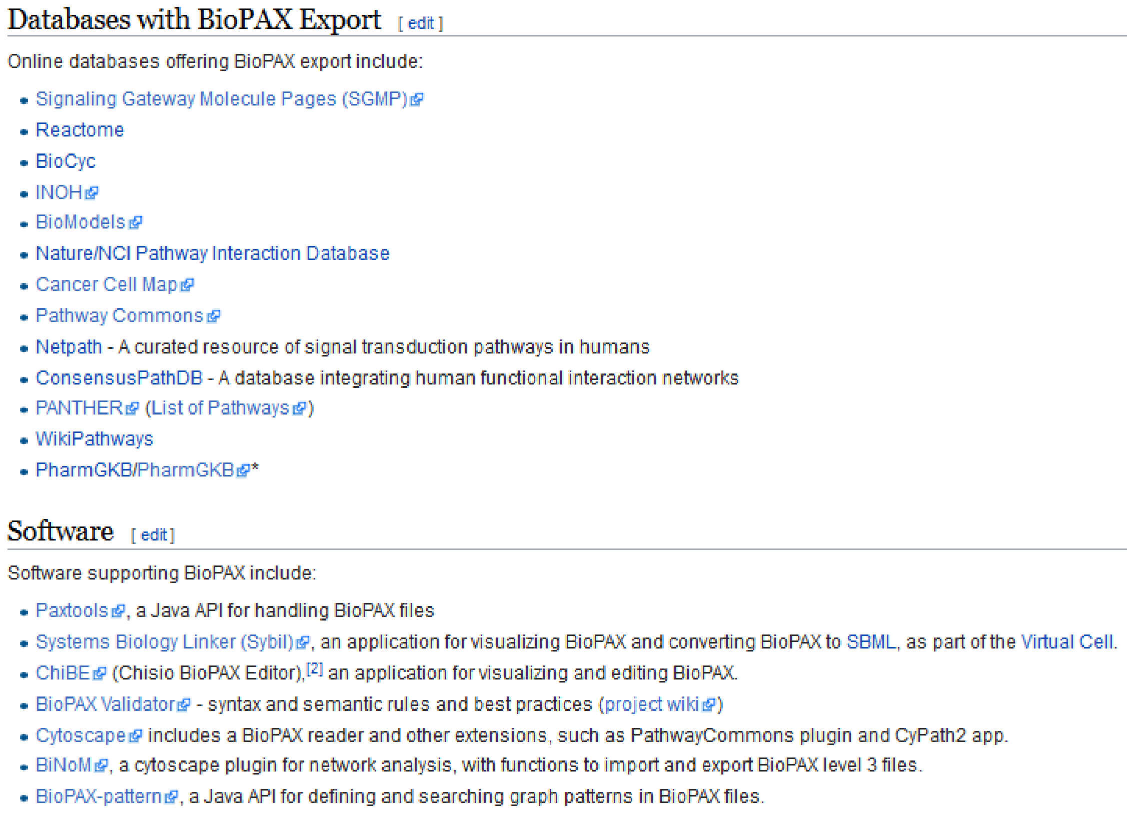
wikipathways.org/index.php/Pathway:WP179

Cell Cycle (Homo sapiens)

Be aware of... Pathway file formats

- Simple graphical file (png, jpeg, etc)
- SBML (Systems Biology Markup Language): Popular in Systems Biology (mathematical models of pathways). Databases of models such as “BioModels”.
- BioPax (Biological Pathway Exchange).

You will need tools that can read the pathway format you choose. Many graphical tools can read SBML and BioPax files.



Databases with BioPAX Export [edit]

Online databases offering BioPAX export include:

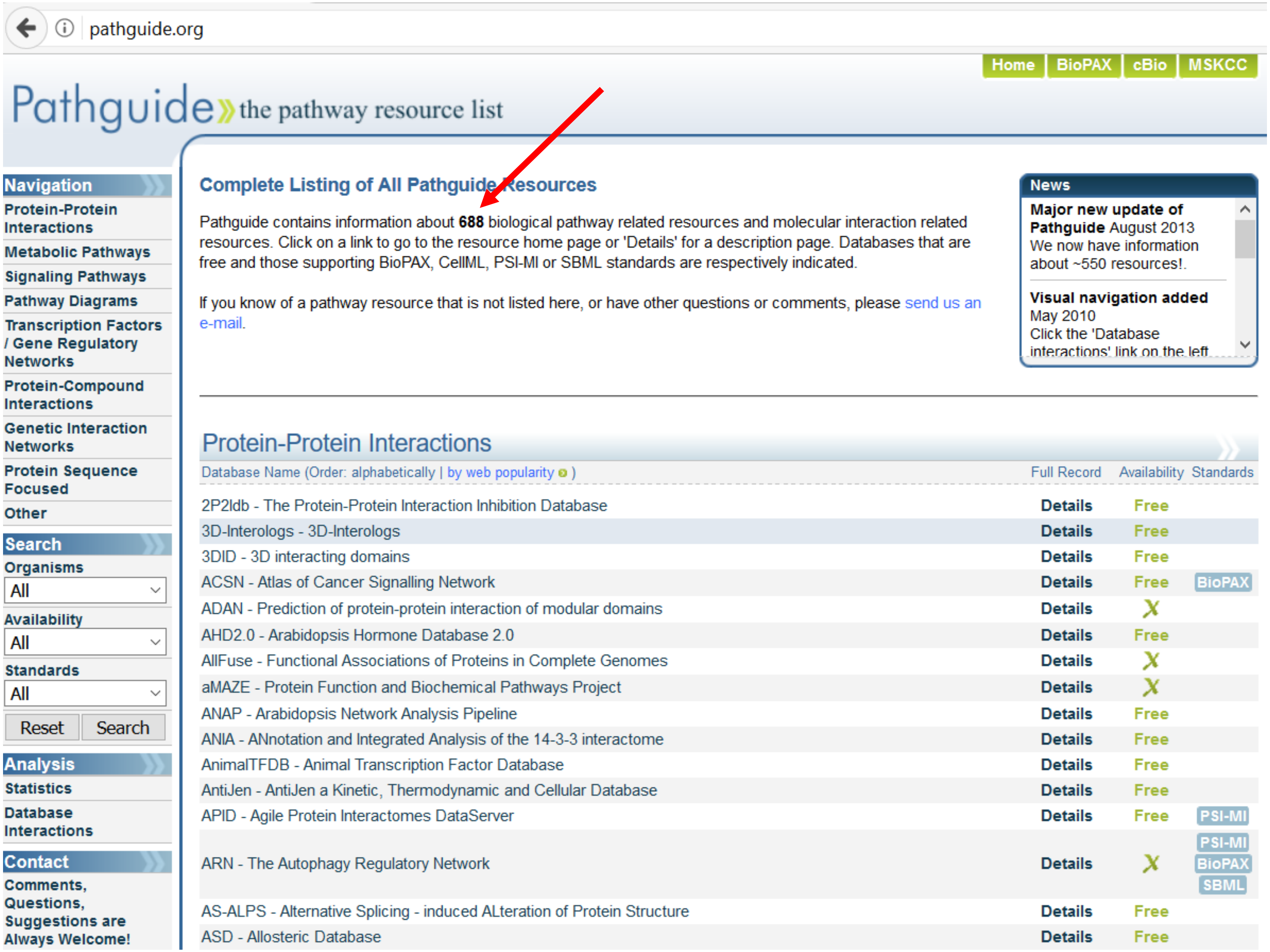
- Signaling Gateway Molecule Pages (SGMP)
- Reactome
- BioCyc
- INOH
- BioModels
- Nature/NCI Pathway Interaction Database
- Cancer Cell Map
- Pathway Commons
- Netpath - A curated resource of signal transduction pathways in humans
- ConsensusPathDB - A database integrating human functional interaction networks
- PANTHER (List of Pathways)
- WikiPathways
- PharmGKB/PharmGKB*

Software [edit]

Software supporting BioPAX include:

- Paxtools, a Java API for handling BioPAX files
- Systems Biology Linker (Sybil), an application for visualizing BioPAX and converting BioPAX to SBML, as part of the Virtual Cell.
- ChiBE (Chisio BioPAX Editor),[2] an application for visualizing and editing BioPAX.
- BioPAX Validator - syntax and semantic rules and best practices (project wiki)
- Cytoscape includes a BioPAX reader and other extensions, such as PathwayCommons plugin and CyPath2 app.
- BiNoM, a cytoscape plugin for network analysis, with functions to import and export BioPAX level 3 files.
- BioPAX-pattern, a Java API for defining and searching graph patterns in BioPAX files.

How many Pathway databases are out there...?



pathguide.org

Pathguide » the pathway resource list

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: All
Availability: All
Standards: All

Analysis
- Statistics
- Database Interactions

Contact

Comments, Questions, Suggestions are Always Welcome!

Complete Listing of All Pathguide Resources

Pathguide contains information about **688** biological pathway related resources and molecular interaction related 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

Major new update of Pathguide August 2013
We now have information about ~550 resources!.

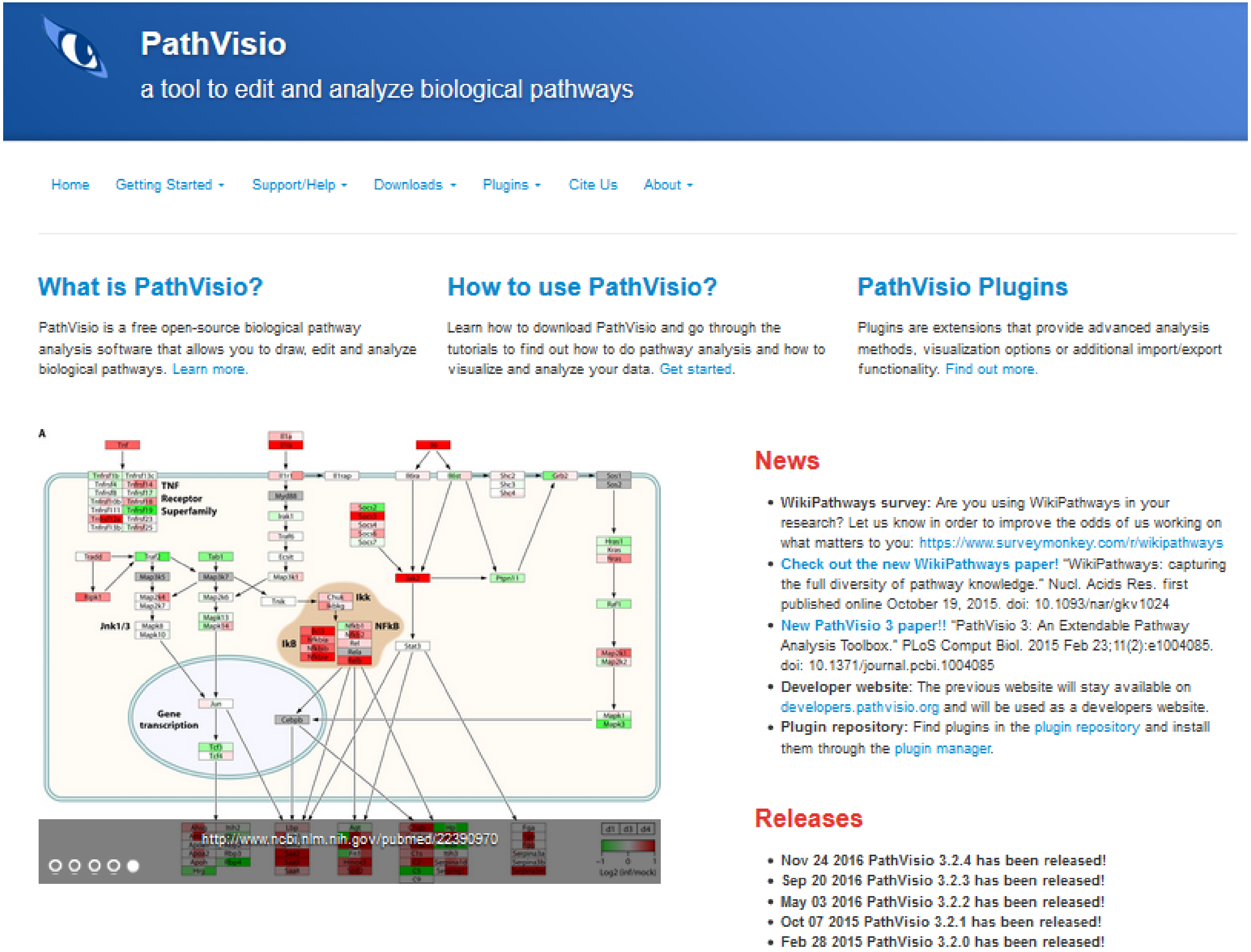
Visual navigation added May 2010
Click the 'Database interactions' link on the left

Protein-Protein Interactions

Database Name (Order: alphabetically | by web popularity)	Full Record	Availability	Standards
2P2Idb - The Protein-Protein Interaction Inhibition Database	Details	Free	
3D-Interologs - 3D-Interologs	Details	Free	
3DID - 3D interacting domains	Details	Free	
ACSN - Atlas of Cancer Signalling Network	Details	Free	BioPAX
ADAN - Prediction of protein-protein interaction of modular domains	Details	X	
AHD2.0 - Arabidopsis Hormone Database 2.0	Details	Free	
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	X	
aMAZE - Protein Function and Biochemical Pathways Project	Details	X	
ANAP - Arabidopsis Network Analysis Pipeline	Details	Free	
ANIA - ANnotation and Integrated Analysis of the 14-3-3 interactome	Details	Free	
AnimalTFDB - Animal Transcription Factor Database	Details	Free	
AntiJen - AntiJen a Kinetic, Thermodynamic and Cellular Database	Details	Free	
APID - Agile Protein Interactomes DataServer	Details	Free	PSI-MI
ARN - The Autophagy Regulatory Network	Details	X	PSI-MI BioPAX SBML
AS-ALPS - Alternative Splicing - induced ALteration of Protein Structure	Details	Free	
ASD - Allosteric Database	Details	Free	

1.3. Pathway visualization

Pathway visualization: PathVisio



PathVisio

a tool to edit and analyze biological pathways

What is PathVisio?

PathVisio is a free open-source biological pathway analysis software that allows you to draw, edit and analyze biological pathways. Learn more.

How to use PathVisio?

Learn how to download PathVisio and go through the tutorials to find out how to do pathway analysis and how to visualize and analyze your data. Get started.

PathVisio Plugins

Plugins are extensions that provide advanced analysis methods, visualization options or additional import/export functionality. Find out more.

http://www.ncbi.nlm.nih.gov/pubmed/22390970

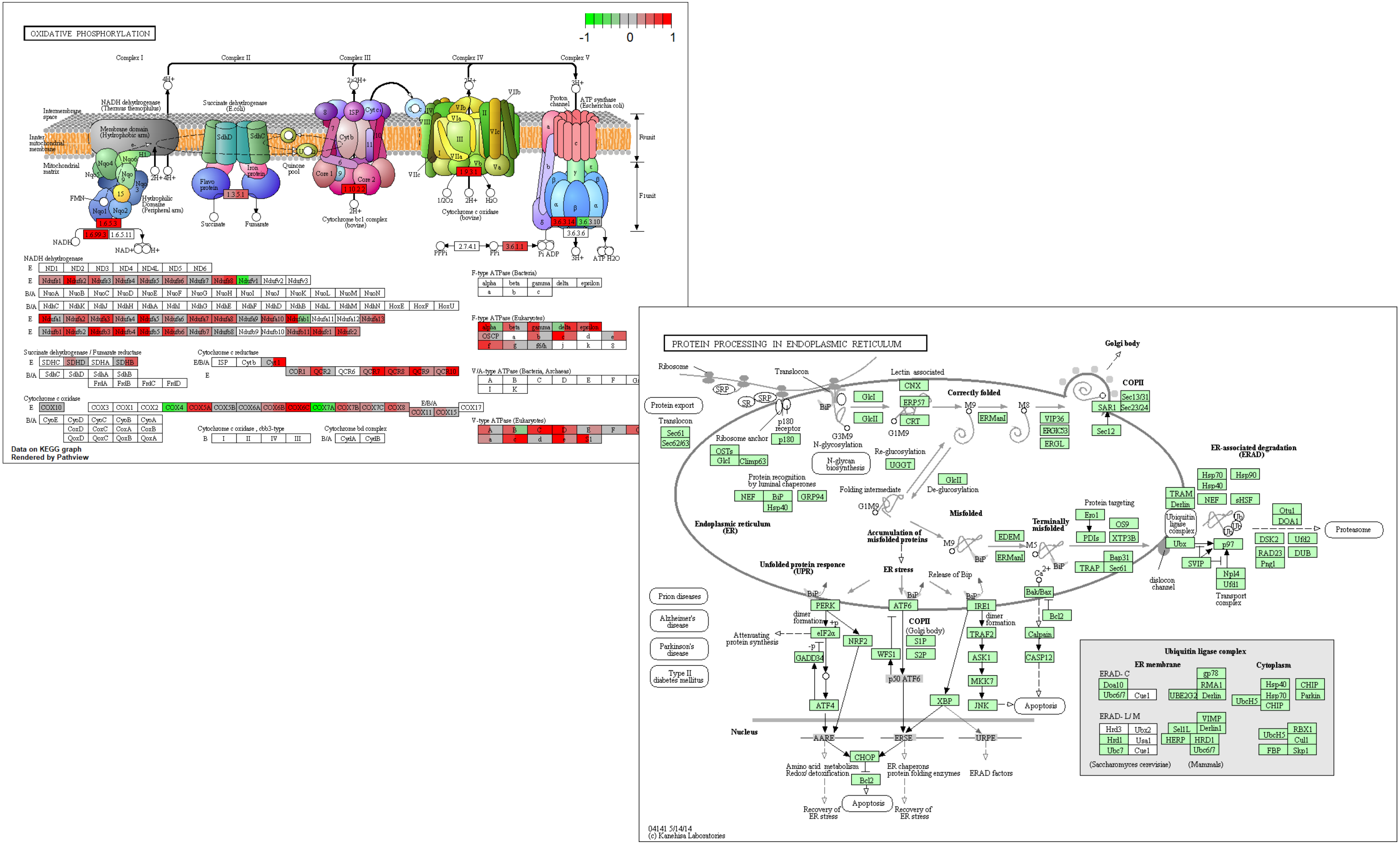
News

- **WikiPathways survey:** Are you using WikiPathways in your research? Let us know in order to improve the odds of us working on what matters to you: https://www.surveymonkey.com/r/wikipathways
- **Check out the new WikiPathways paper!** "WikiPathways: capturing the full diversity of pathway knowledge." Nucl. Acids Res. first published online October 19, 2015. doi: 10.1093/nar/gkv1024
- **New PathVisio 3 paper!!** "PathVisio 3: An Extendable Pathway Analysis Toolbox." PLoS Comput Biol. 2015 Feb 23;11(2):e1004085. doi: 10.1371/journal.pcbi.1004085
- **Developer website:** The previous website will stay available on developers.pathvisio.org and will be used as a developers website.
- **Plugin repository:** Find plugins in the plugin repository and install them through the plugin manager.

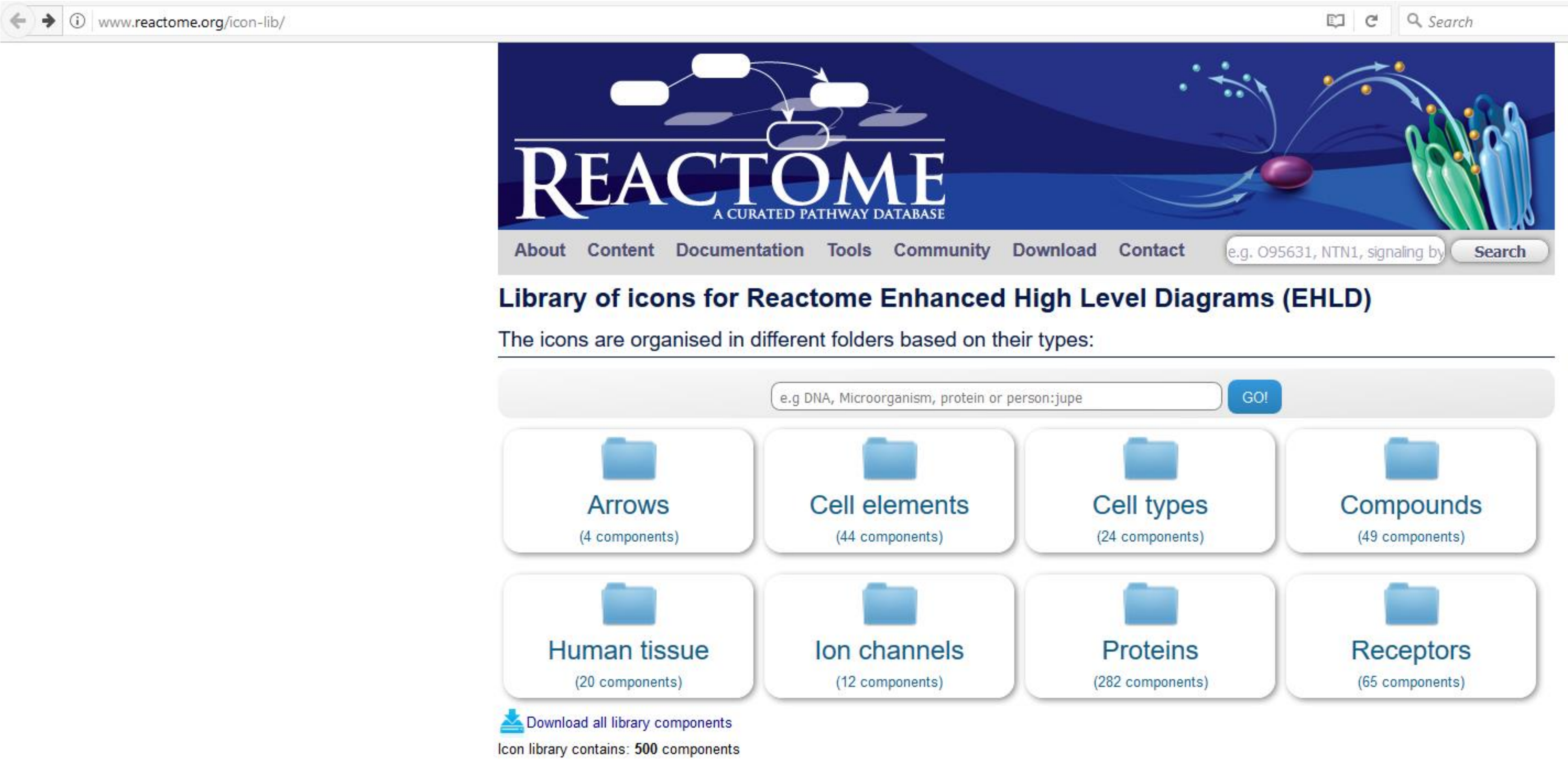
Releases

- **Nov 24 2016 PathVisio 3.2.4 has been released!**
- **Sep 20 2016 PathVisio 3.2.3 has been released!**
- **May 03 2016 PathVisio 3.2.2 has been released!**
- **Oct 07 2015 PathVisio 3.2.1 has been released!**
- **Feb 28 2015 PathVisio 3.2.0 has been released!**

Pathway visualization: R / pathview



Pathway visualization: Reactome Library of Icons



www.reactome.org/icon-lib/
Search
REACTOME
A CURATED PATHWAY DATABASE
About
Content
Documentation
Tools
Community
Download
Contact
e.g. O95631, NTN1, signaling by
Search

Library of icons for Reactome Enhanced High Level Diagrams (EHLD)

The icons are organised in different folders based on their types:

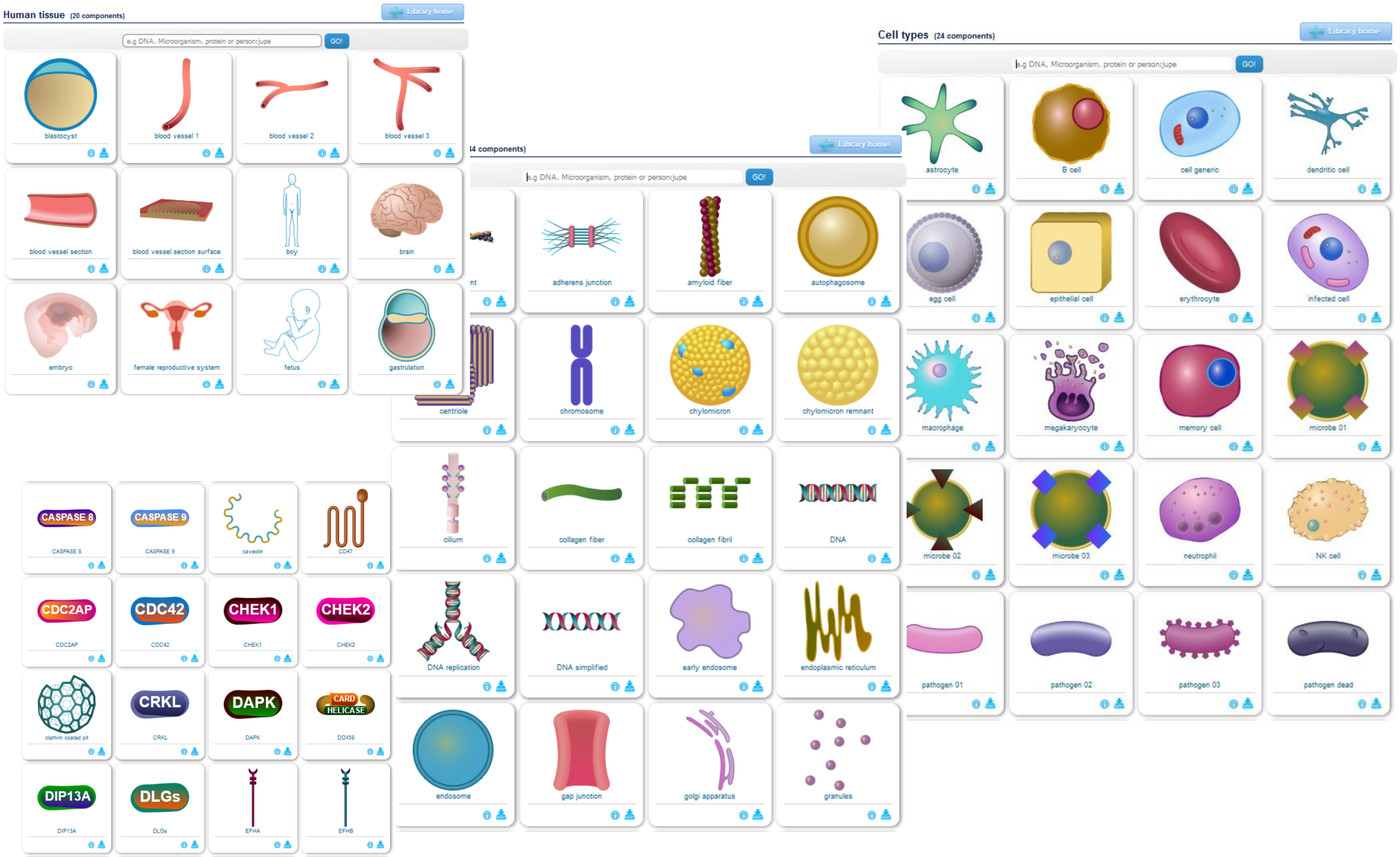
e.g DNA, Microorganism, protein or person:jupe
GO!

Arrows (4 components)	Cell elements (44 components)	Cell types (24 components)	Compounds (49 components)
Human tissue (20 components)	Ion channels (12 components)	Proteins (282 components)	Receptors (65 components)

Download all library components

Icon library contains: **500** components

Pathway visualization: Reactome Library of Icons



1.4. The Gene Ontology (GO)

What is the Gene Ontology (GO)?

CAN YOU PLEASE SUMMARIZE ALL CONCEPTS IN THIS BIOLOGY BOOK AND TELL ME HOW THEY RELATE TO EACH OTHER?

SURE! BASICALLY, THERE ARE 3 BASIC THINGS HERE: BIOLOGICAL PROCESSES, MOLECULAR FUNCTIONS, AND CELLULAR COMPONENTS. NOW, THE BIOLOGICAL PROCESSES CAN BE DIVIDED INTO…

WOW! HE ORGANIZED ALL BIOLOGICAL KNOWLEDGE IN AN **ONTOLOGY!!**

https://www.freepik.com/

What is the Gene Ontology (GO)?

ALSO… CAN YOU FIND A WAY TO TELL ME ALL THE BIOLOGY CONCEPTS RELATED TO A GIVEN GENE?

SURE! WE BUILT THIS DATABASE CALLED “GO” WHERE EVERY GENE IS RELATED TO EVERY CONCEPT IN OUR ONTOLOGY

WOW! HIS ONTOLOGY IS **ANNOTATED!!**

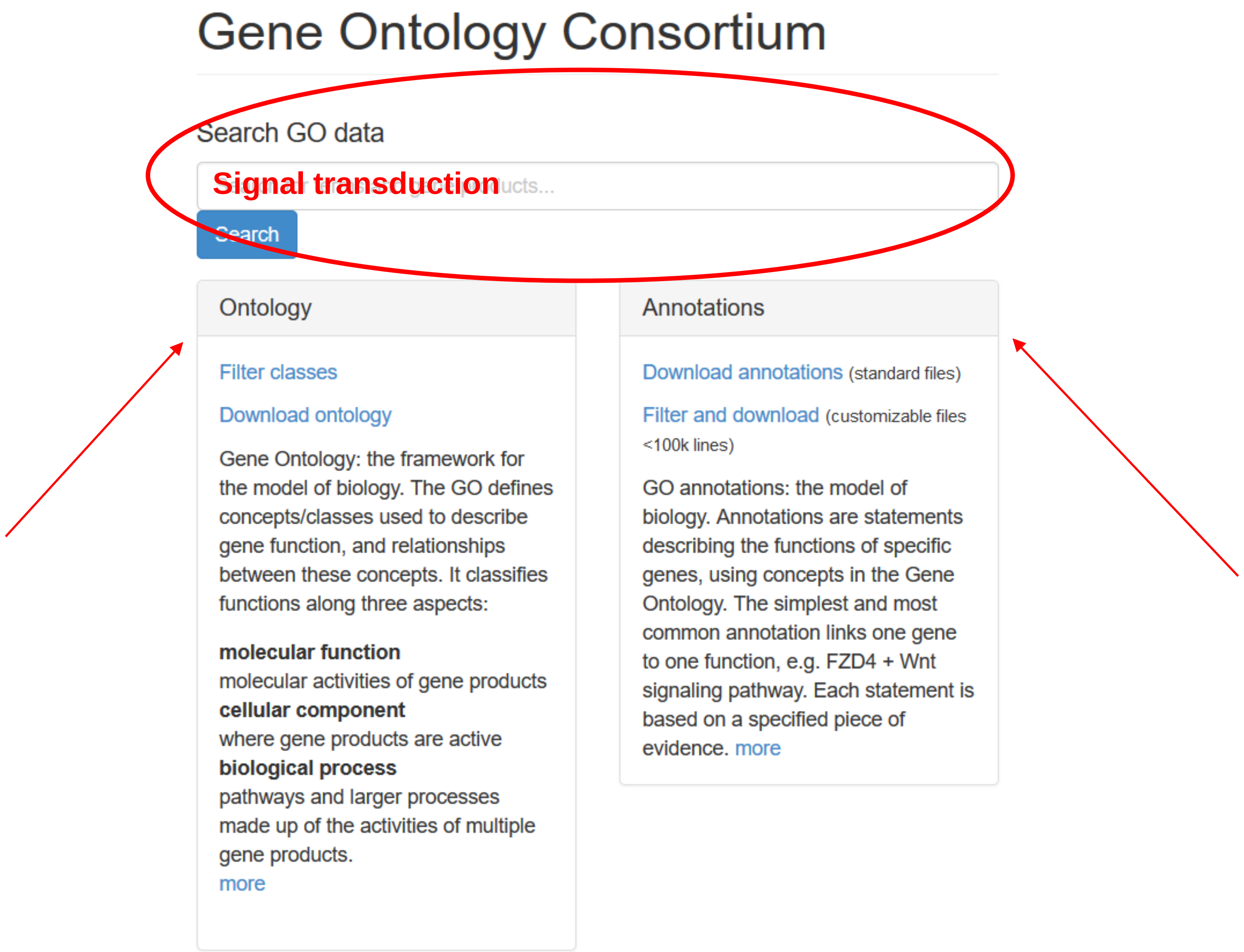
https://www.freepik.com/

What is the Gene Ontology (GO)?

- Set of words / phrases (called GO terms) which are related to genes. For example: “protein kinase”, “glycolysis”, “nucleus”.
- It is a Dictionary: Term definitions
- It is an Ontology: A formal system for describing knowledge
- It is Annotated: Genes linked to GO terms

Gene Ontology Consortium

The Gene Ontology (GO)



Gene Ontology Consortium

Search GO data

Signal transduction

Search

Ontology

Filter classes

Download ontology

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function
molecular activities of gene products
cellular component
where gene products are active
biological process
pathways and larger processes made up of the activities of multiple gene products.
more

Annotations

Download annotations (standard files)

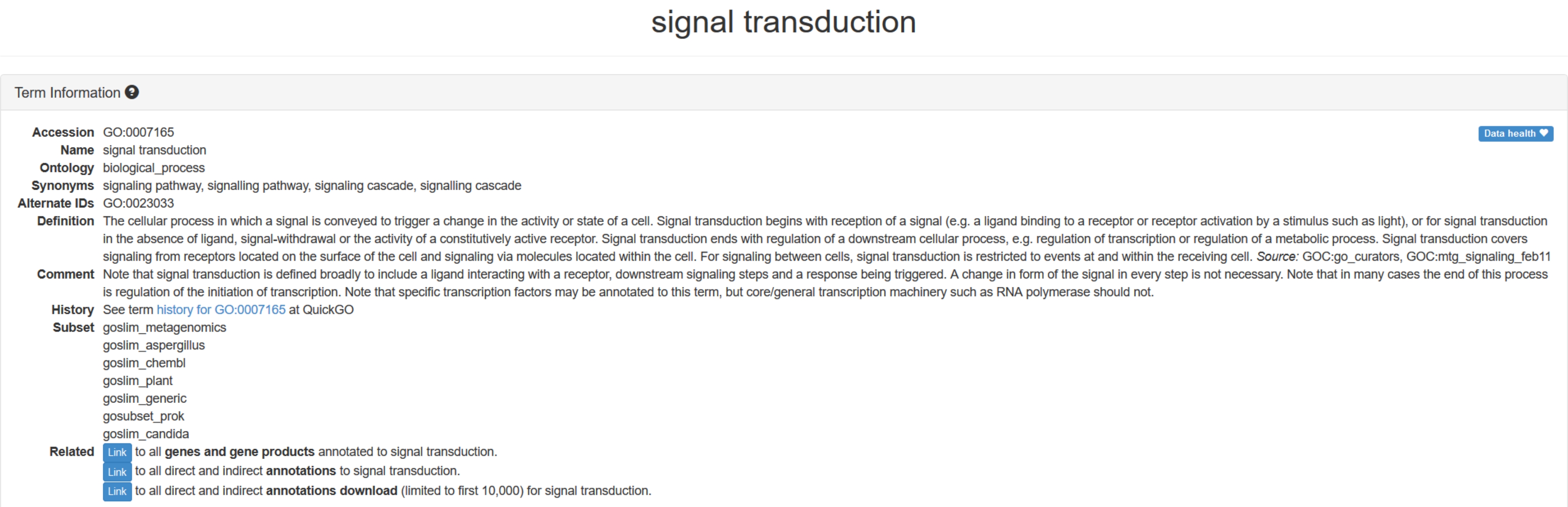
Filter and download (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. more

Gene Ontology Consortium

The Gene Ontology (GO)

Description (Name, Ontology, GO Term, Synonym, Definitions):



signal transduction

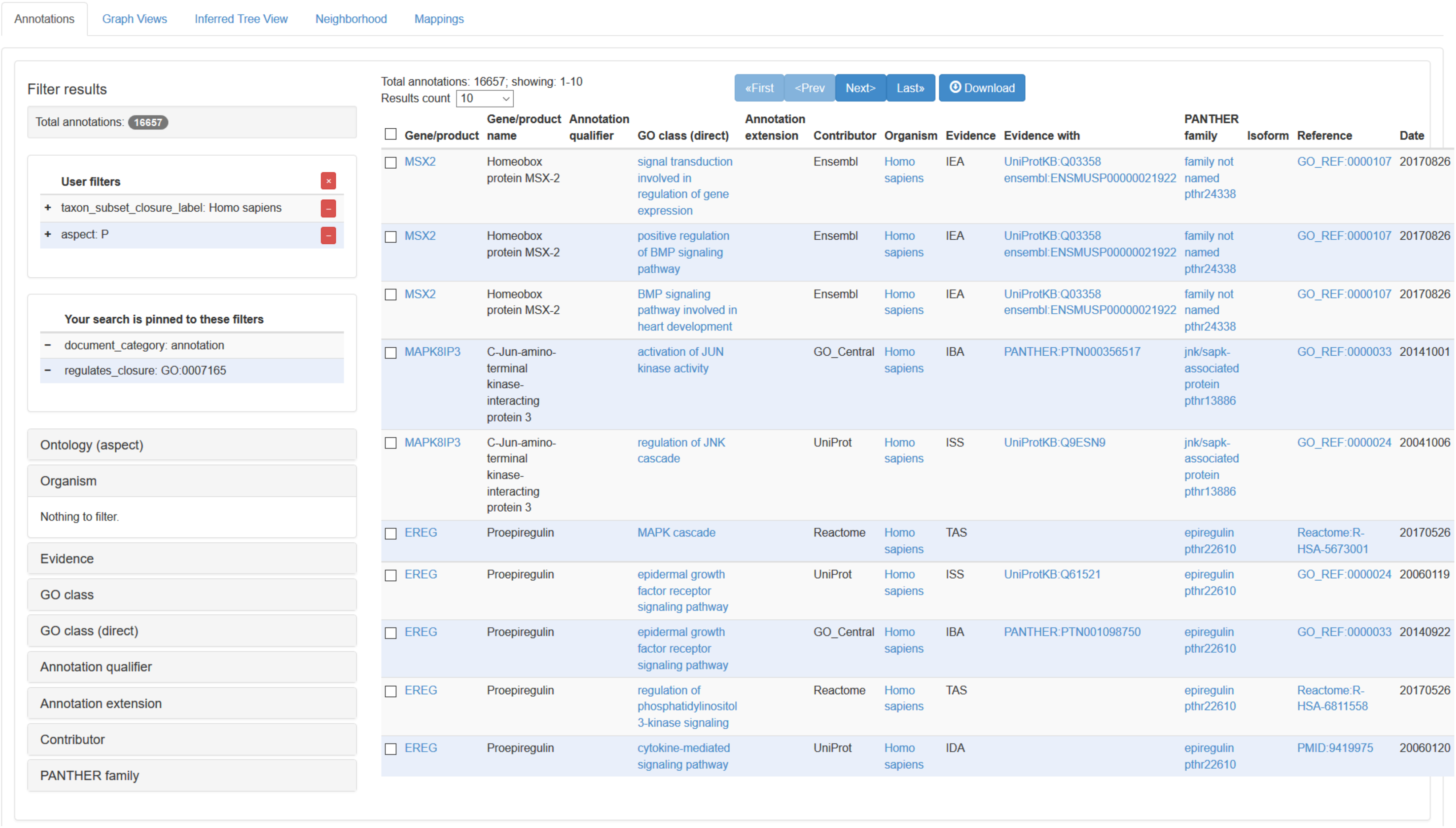
Term Information

Accession GO:0007165
Name signal transduction
Ontology biological_process
Synonyms signaling pathway, signalling pathway, signaling cascade, signalling cascade
Alternate IDs GO:0023033
Definition The cellular process in which a signal is conveyed to trigger a change in the activity or state of a cell. Signal transduction begins with reception of a signal (e.g. a ligand binding to a receptor or receptor activation by a stimulus such as light), or for signal transduction in the absence of ligand, signal-withdrawal or the activity of a constitutively active receptor. Signal transduction ends with regulation of a downstream cellular process, e.g. regulation of transcription or regulation of a metabolic process. Signal transduction covers signaling from receptors located on the surface of the cell and signaling via molecules located within the cell. For signaling between cells, signal transduction is restricted to events at and within the receiving cell. *Source:* GOC:go_curators, GOC:mtg_signaling_feb11
Comment Note that signal transduction is defined broadly to include a ligand interacting with a receptor, downstream signaling steps and a response being triggered. A change in form of the signal in every step is not necessary. Note that in many cases the end of this process is regulation of the initiation of transcription. Note that specific transcription factors may be annotated to this term, but core/general transcription machinery such as RNA polymerase should not.
History See term history for GO:0007165 at QuickGO
Subset goslim_metagenomics
goslim_aspergillus
goslim_chembl
goslim_plant
goslim_generic
gosubset_prok
goslim_candida
Related Link to all **genes and gene products** annotated to signal transduction.
Link to all direct and indirect **annotations** to signal transduction.
Link to all direct and indirect **annotations download** (limited to first 10,000) for signal transduction.

Gene Ontology Consortium

The Gene Ontology (GO)

Annotations:



Filter results

Total annotations: 16657

User filters ×

- + taxon_subset_closure_label: Homo sapiens –
- + aspect: P –

Your search is pinned to these filters

- – document_category: annotation
- – regulates_closure: GO:0007165

Ontology (aspect)

Organism

Nothing to filter.

Evidence

GO class

GO class (direct)

Annotation qualifier

Annotation extension

Contributor

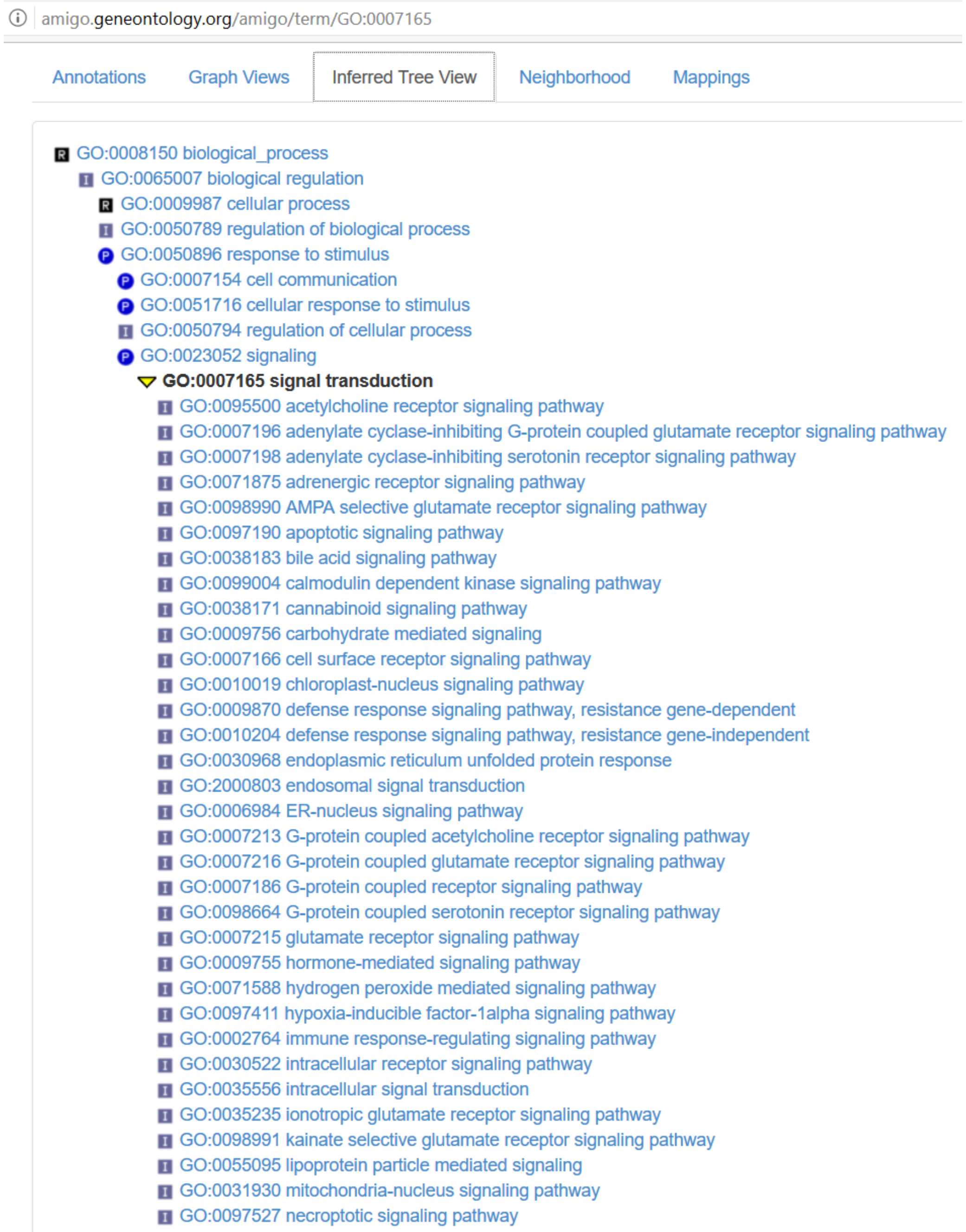
PANTHER family

Total annotations: 16657; showing: 1-10
Results count 10

□	Gene/product	Gene/product name	Annotation qualifier	GO class (direct)	Annotation extension	Contributor	Organism	Evidence	Evidence with	PANTHER family	Isoform	Reference	Date
□	MSX2	Homeobox protein MSX-2		signal transduction involved in regulation of gene expression		Ensembl	Homo sapiens	IEA	UniProtKB:Q03358 ensembl:ENSMUSP00000021922	family not named pthr24338		GO_REF:0000107	20170826
□	MSX2	Homeobox protein MSX-2		positive regulation of BMP signaling pathway		Ensembl	Homo sapiens	IEA	UniProtKB:Q03358 ensembl:ENSMUSP00000021922	family not named pthr24338		GO_REF:0000107	20170826
□	MSX2	Homeobox protein MSX-2		BMP signaling pathway involved in heart development		Ensembl	Homo sapiens	IEA	UniProtKB:Q03358 ensembl:ENSMUSP00000021922	family not named pthr24338		GO_REF:0000107	20170826
□	MAPK8IP3	C-Jun-amino-terminal kinase-interacting protein 3		activation of JUN kinase activity		GO_Central	Homo sapiens	IBA	PANTHER:PTN000356517	jnk/sapk-associated protein pthr13886		GO_REF:0000033	20141001
□	MAPK8IP3	C-Jun-amino-terminal kinase-interacting protein 3		regulation of JNK cascade		UniProt	Homo sapiens	ISS	UniProtKB:Q9ESN9	jnk/sapk-associated protein pthr13886		GO_REF:0000024	20041006
□	EREG	Proepiregulin		MAPK cascade		Reactome	Homo sapiens	TAS		epiregulin pthr22610		Reactome:R-HSA-5673001	20170526
□	EREG	Proepiregulin		epidermal growth factor receptor signaling pathway		UniProt	Homo sapiens	ISS	UniProtKB:Q61521	epiregulin pthr22610		GO_REF:0000024	20060119
□	EREG	Proepiregulin		epidermal growth factor receptor signaling pathway		GO_Central	Homo sapiens	IBA	PANTHER:PTN001098750	epiregulin pthr22610		GO_REF:0000033	20140922
□	EREG	Proepiregulin		regulation of phosphatidylinositol 3-kinase signaling		Reactome	Homo sapiens	TAS		epiregulin pthr22610		Reactome:R-HSA-6811558	20170526
□	EREG	Proepiregulin		cytokine-mediated signaling pathway		UniProt	Homo sapiens	IDA		epiregulin pthr22610		PMID:9419975	20060120

Ontology tree:

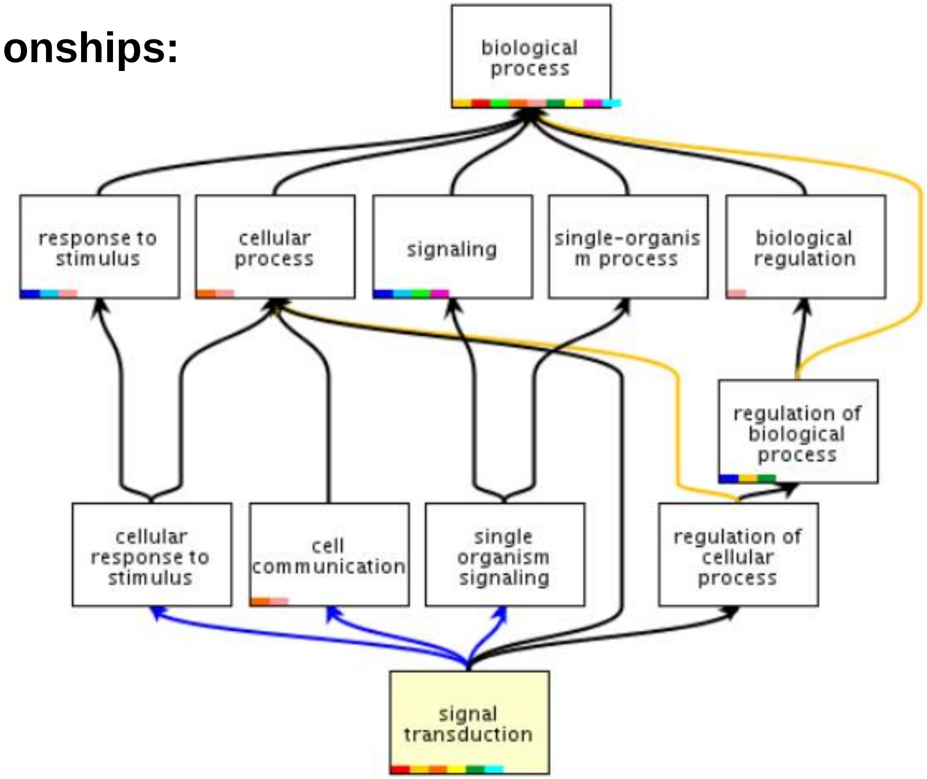
Parents and chlidren

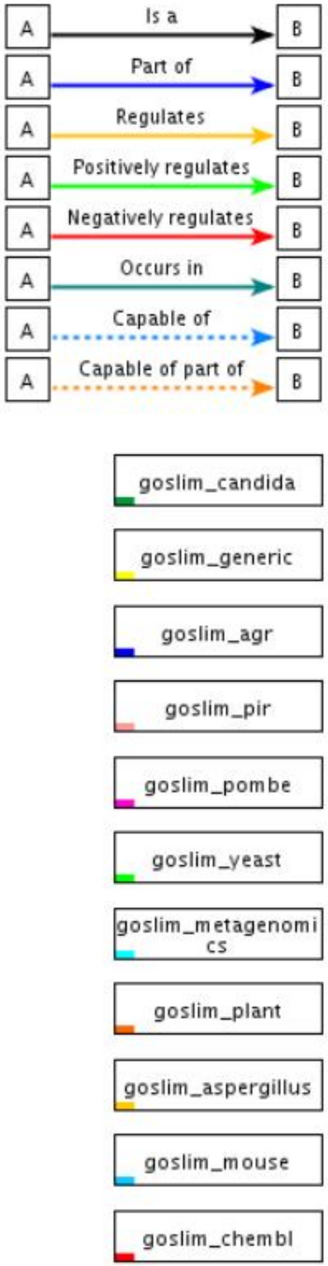


amigo.geneontology.org/amigo/term/GO:0007165

- R GO:0008150 biological_process
  - I GO:0065007 biological regulation
    - R GO:0009987 cellular process
    - I GO:0050789 regulation of biological process
    - P GO:0050896 response to stimulus
      - P GO:0007154 cell communication
      - P GO:0051716 cellular response to stimulus
      - I GO:0050794 regulation of cellular process
      - P GO:0023052 signaling
        - ▽ **GO:0007165 signal transduction**
          - I GO:0095500 acetylcholine receptor signaling pathway
          - I GO:0007196 adenylate cyclase-inhibiting G-protein coupled glutamate receptor signaling pathway
          - I GO:0007198 adenylate cyclase-inhibiting serotonin receptor signaling pathway
          - I GO:0071875 adrenergic receptor signaling pathway
          - I GO:0098990 AMPA selective glutamate receptor signaling pathway
          - I GO:0097190 apoptotic signaling pathway
          - I GO:0038183 bile acid signaling pathway
          - I GO:0099004 calmodulin dependent kinase signaling pathway
          - I GO:0038171 cannabinoid signaling pathway
          - I GO:0009756 carbohydrate mediated signaling
          - I GO:0007166 cell surface receptor signaling pathway
          - I GO:0010019 chloroplast-nucleus signaling pathway
          - I GO:0009870 defense response signaling pathway, resistance gene-dependent
          - I GO:0010204 defense response signaling pathway, resistance gene-independent
          - I GO:0030968 endoplasmic reticulum unfolded protein response
          - I GO:2000803 endosomal signal transduction
          - I GO:0006984 ER-nucleus signaling pathway
          - I GO:0007213 G-protein coupled acetylcholine receptor signaling pathway
          - I GO:0007216 G-protein coupled glutamate receptor signaling pathway
          - I GO:0007186 G-protein coupled receptor signaling pathway
          - I GO:0098664 G-protein coupled serotonin receptor signaling pathway
          - I GO:0007215 glutamate receptor signaling pathway
          - I GO:0009755 hormone-mediated signaling pathway
          - I GO:0071588 hydrogen peroxide mediated signaling pathway
          - I GO:0097411 hypoxia-inducible factor-1alpha signaling pathway
          - I GO:0002764 immune response-regulating signaling pathway
          - I GO:0030522 intracellular receptor signaling pathway
          - I GO:0035556 intracellular signal transduction
          - I GO:0035235 ionotropic glutamate receptor signaling pathway
          - I GO:0098991 kainate selective glutamate receptor signaling pathway
          - I GO:0055095 lipoprotein particle mediated signaling
          - I GO:0031930 mitochondria-nucleus signaling pathway
          - I GO:0097527 necroptotic signaling pathway

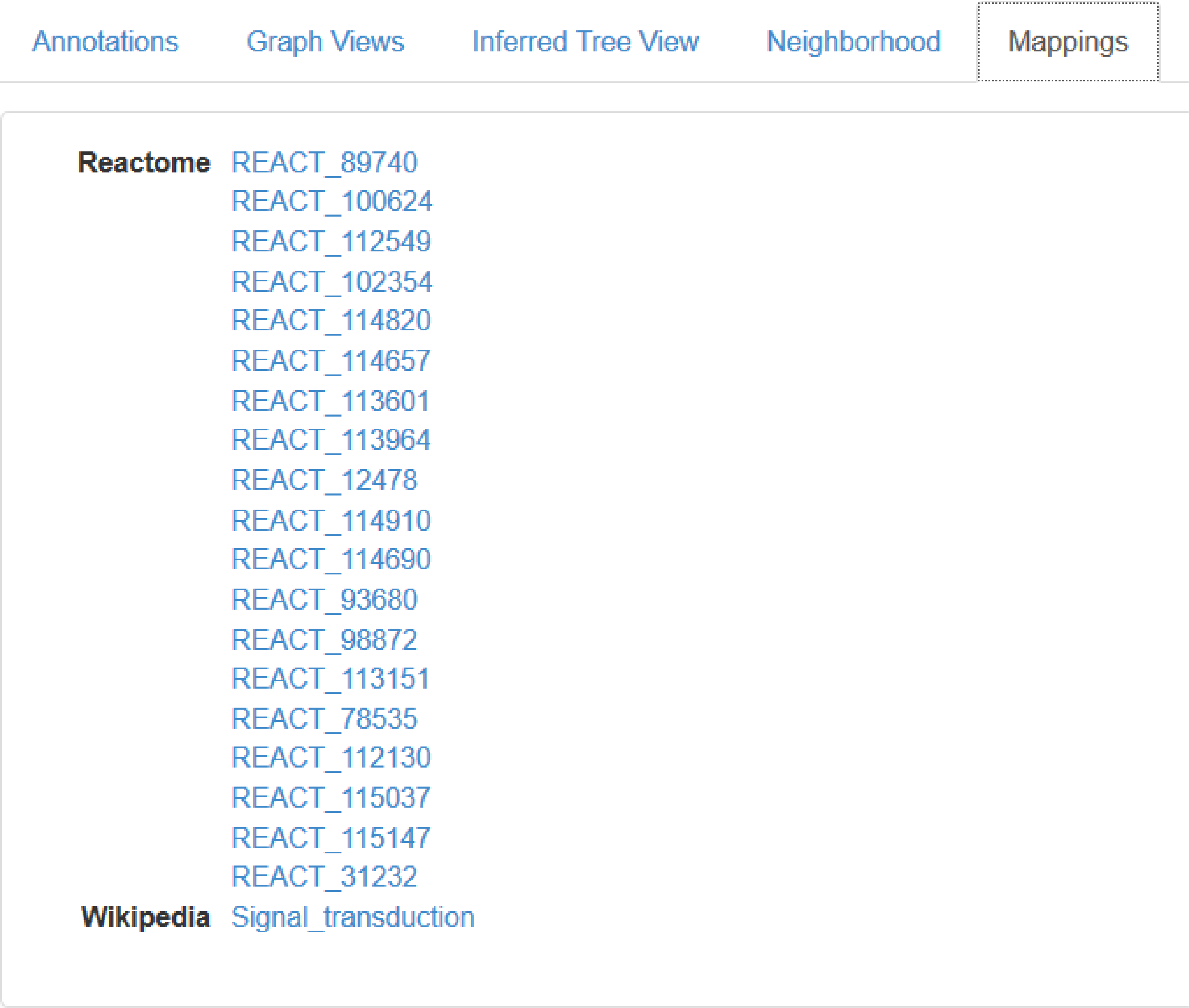
Ontology relationships:





- Terms are related within a hierarchy
- Describes multiple levels of detail of gene function
- Terms can have more than one parent or child

QuickGO - http://www.ebi.ac.uk/QuickGO



Reactome	REACT_89740 REACT_100624 REACT_112549 REACT_102354 REACT_114820 REACT_114657 REACT_113601 REACT_113964 REACT_12478 REACT_114910 REACT_114690 REACT_93680 REACT_98872 REACT_113151 REACT_78535 REACT_112130 REACT_115037 REACT_115147 REACT_31232
Wikipedia	Signal_transduction

GO Terms and GO Annotations

- GO terms are added by editors at EBI
- Some terms may be added by request

- Genes are associated with GO terms either by trained curators or created automatically (without human review)
- Multiple annotations per gene
- Manual annotation is considered of higher quality but it is time-consuming.
- Electronic annotation may have variable quality.

Evidence Types

- Experimental Evidence Codes
  - EXP: Inferred from Experiment
  - IDA: Inferred from Direct Assay
  - IPI: Inferred from Physical Interaction
  - IMP: Inferred from Mutant Phenotype
  - IGI: Inferred from Genetic Interaction
  - IEP: Inferred from Expression Pattern

- Computational Analysis Evidence Codes
  - ISS: Inferred from Sequence or Structural Similarity
  - ISO: Inferred from Sequence Orthology
  - ISA: Inferred from Sequence Alignment
  - ISM: Inferred from Sequence Model
  - IGC: Inferred from Genomic Context
  - RCA: inferred from Reviewed Computational Analysis

- Author Statement Evidence Codes
  - TAS: Traceable Author Statement
  - NAS: Non-traceable Author Statement
- Curator Statement Evidence Codes
  - IC: Inferred by Curator
  - ND: No biological Data available

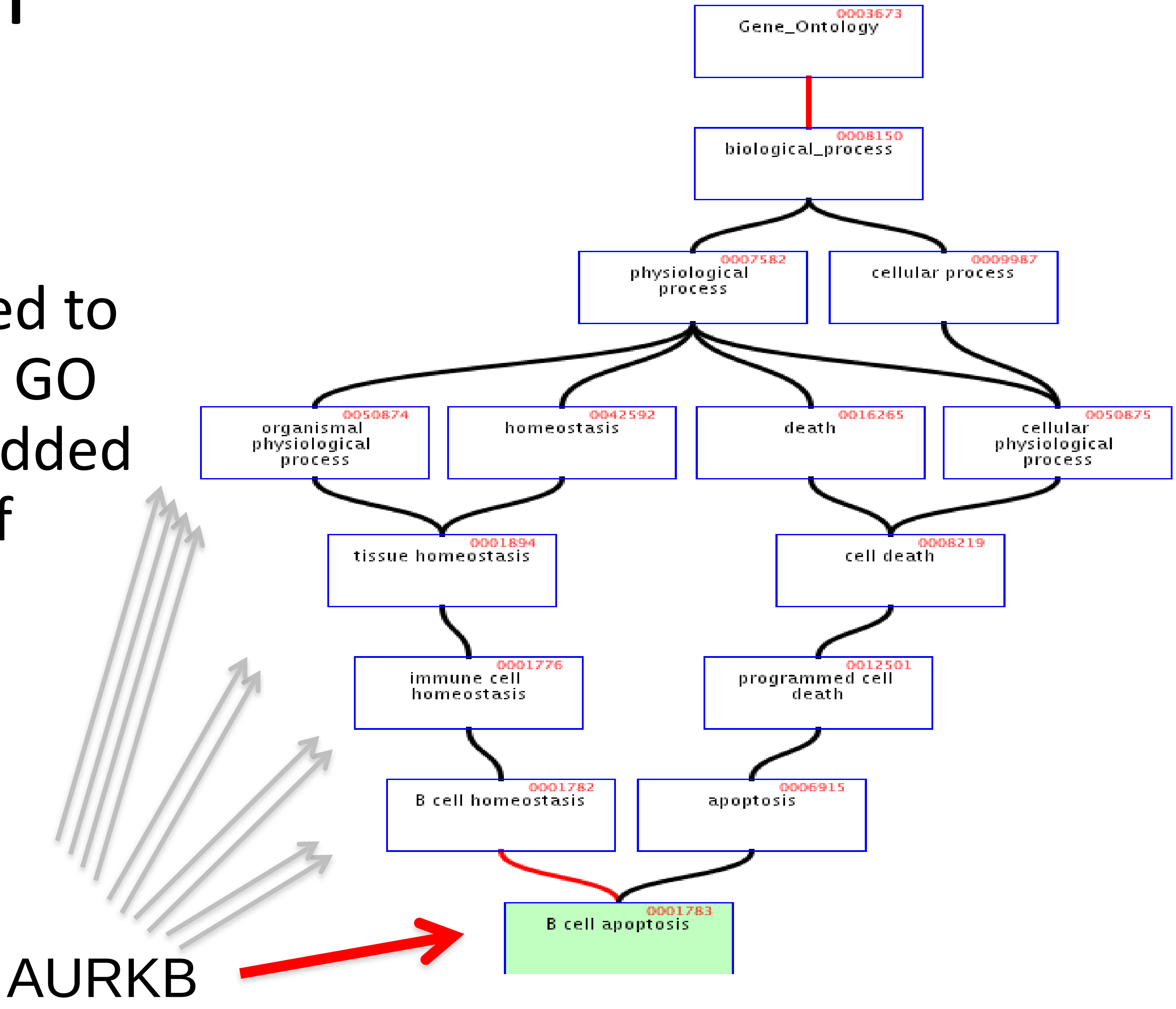
- IEA: Inferred from electronic annotation

Key point: be aware of annotation origin

http://www.geneontology.org/GO.evidence.shtml

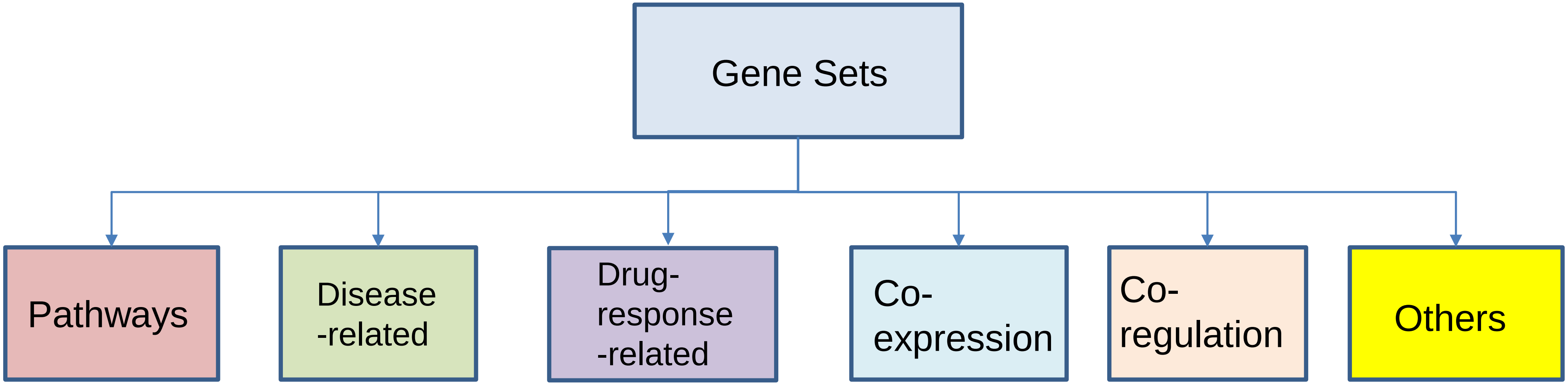
Hierarchical annotation

- Genes annotated to specific term in GO automatically added to *all parents* of that term



1.5. «Gene set» databases

From pathways to «gene sets»



From pathway databases to “gene set” databases, such as **GeneSetDB** (Araki, 2012) and **MSigDB** (Broad Institute), which include pathways, phenotypes, GO, and others.

GeneSetDB

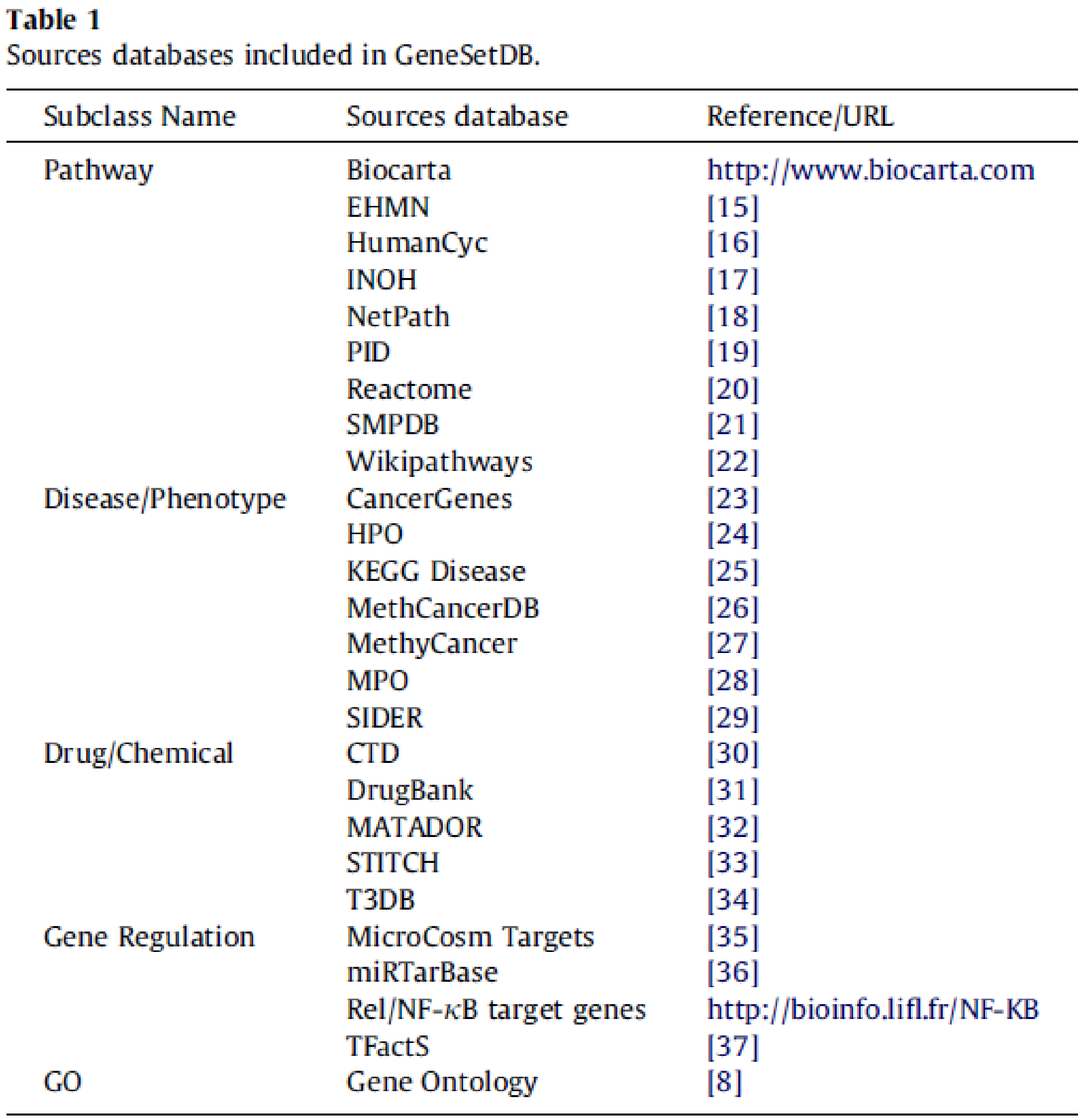
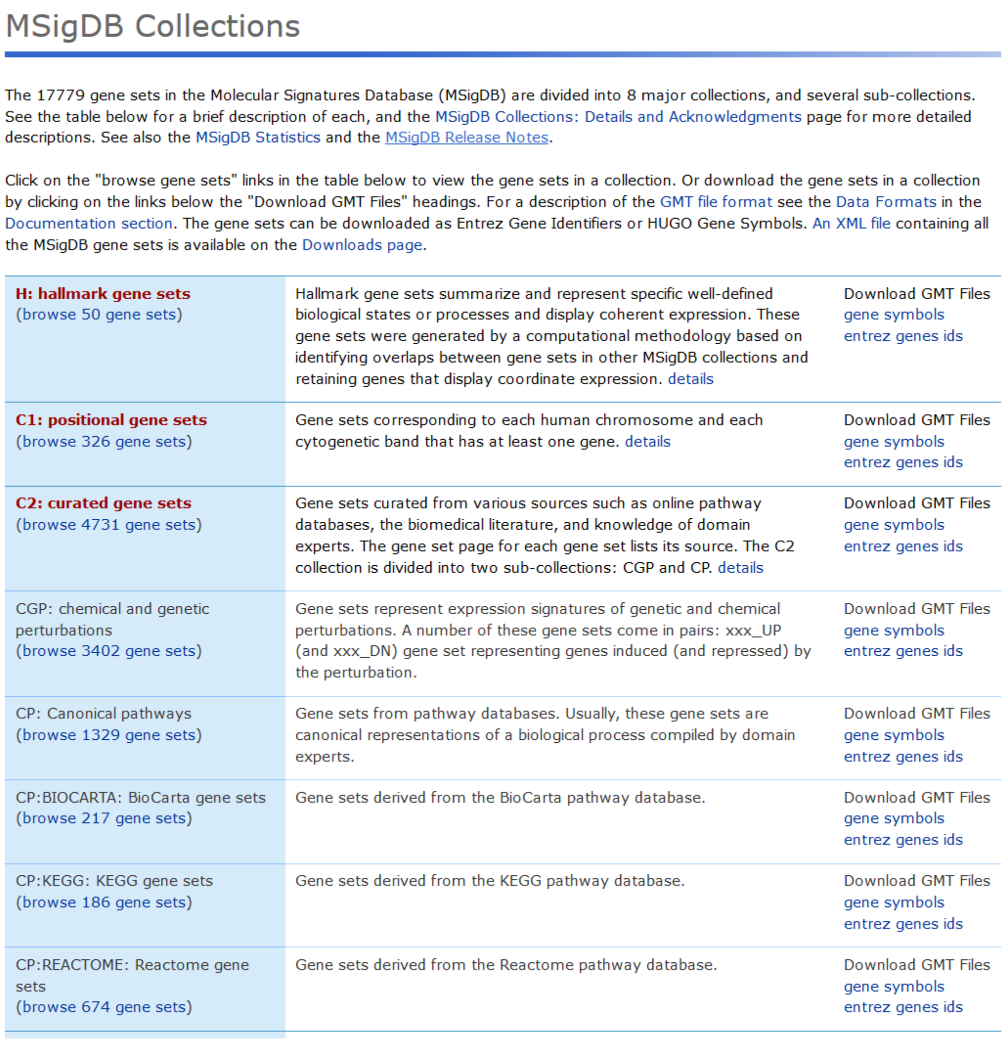


Table 1
Sources databases included in GeneSetDB.

Subclass Name	Sources database	Reference/URL
Pathway	Biocarta	http://www.biocarta.com
	EHMN	[15]
	HumanCyc	[16]
	INOH	[17]
	NetPath	[18]
	PID	[19]
	Reactome	[20]
	SMPDB	[21]
	Wikipathways	[22]
Disease/Phenotype	CancerGenes	[23]
	HPO	[24]
	KEGG Disease	[25]
	MethCancerDB	[26]
	MethyCancer	[27]
	MPO	[28]
	SIDER	[29]
Drug/Chemical	CTD	[30]
	DrugBank	[31]
	MATADOR	[32]
	STITCH	[33]
	T3DB	[34]
Gene Regulation	MicroCosm Targets	[35]
	miRTarBase	[36]
	Rel/NF-κB target genes	http://bioinfo.lifl.fr/NF-KB
	TFactS	[37]
GO	Gene Ontology	[8]

MSigDB

MSigDB Molecular Signatures Database

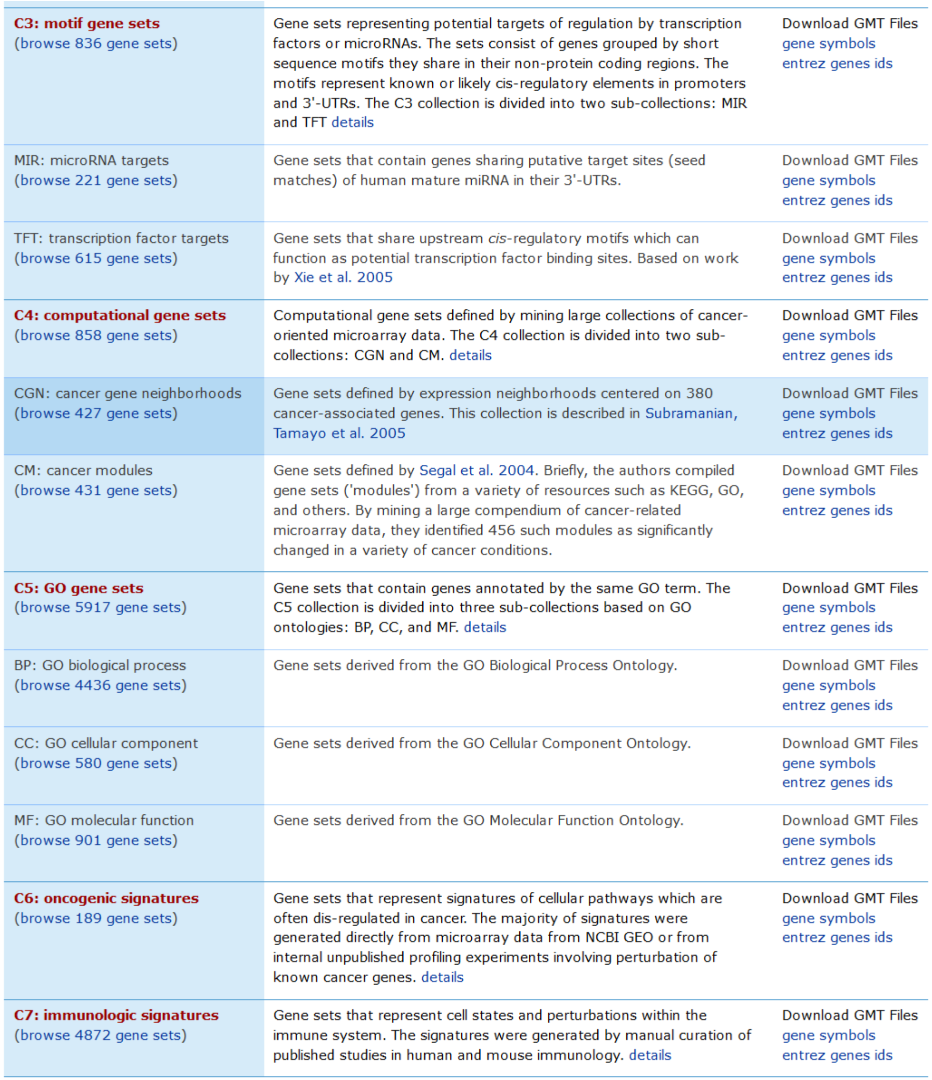


MSigDB Collections

The 17779 gene sets in the Molecular Signatures Database (MSigDB) are divided into 8 major collections, and several sub-collections. See the table below for a brief description of each, and the MSigDB Collections: Details and Acknowledgments page for more detailed descriptions. See also the MSigDB Statistics and the MSigDB Release Notes.

Click on the "browse gene sets" links in the table below to view the gene sets in a collection. Or download the gene sets in a collection by clicking on the links below the "Download GMT Files" headings. For a description of the GMT file format see the Data Formats in the Documentation section. The gene sets can be downloaded as Entrez Gene Identifiers or HUGO Gene Symbols. An XML file containing all the MSigDB gene sets is available on the Downloads page.

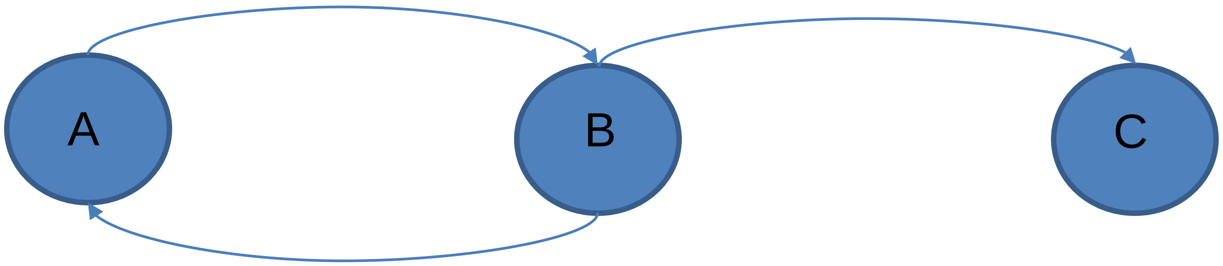
H: hallmark gene sets (browse 50 gene sets)	Hallmark gene sets summarize and represent specific well-defined biological states or processes and display coherent expression. These gene sets were generated by a computational methodology based on identifying overlaps between gene sets in other MSigDB collections and retaining genes that display coordinate expression. details	Download GMT Files gene symbols entrez genes ids
C1: positional gene sets (browse 326 gene sets)	Gene sets corresponding to each human chromosome and each cytogenetic band that has at least one gene. details	Download GMT Files gene symbols entrez genes ids
C2: curated gene sets (browse 4731 gene sets)	Gene sets curated from various sources such as online pathway databases, the biomedical literature, and knowledge of domain experts. The gene set page for each gene set lists its source. The C2 collection is divided into two sub-collections: CGP and CP. details	Download GMT Files gene symbols entrez genes ids
CGP: chemical and genetic perturbations (browse 3402 gene sets)	Gene sets represent expression signatures of genetic and chemical perturbations. A number of these gene sets come in pairs: xxx_UP (and xxx_DN) gene set representing genes induced (and repressed) by the perturbation.	Download GMT Files gene symbols entrez genes ids
CP: Canonical pathways (browse 1329 gene sets)	Gene sets from pathway databases. Usually, these gene sets are canonical representations of a biological process compiled by domain experts.	Download GMT Files gene symbols entrez genes ids
CP:BIOCARTA: BioCarta gene sets (browse 217 gene sets)	Gene sets derived from the BioCarta pathway database.	Download GMT Files gene symbols entrez genes ids
CP:KEGG: KEGG gene sets (browse 186 gene sets)	Gene sets derived from the KEGG pathway database.	Download GMT Files gene symbols entrez genes ids
CP:REACTOME: Reactome gene sets (browse 674 gene sets)	Gene sets derived from the Reactome pathway database.	Download GMT Files gene symbols entrez genes ids



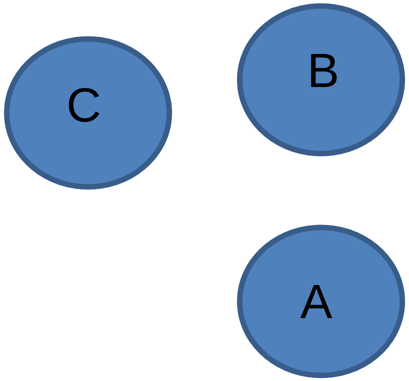
C3: motif gene sets (browse 836 gene sets)	Gene sets representing potential targets of regulation by transcription factors or microRNAs. The sets consist of genes grouped by short sequence motifs they share in their non-protein coding regions. The motifs represent known or likely cis-regulatory elements in promoters and 3'-UTRs. The C3 collection is divided into two sub-collections: MIR and TFT details	Download GMT Files gene symbols entrez genes ids
MIR: microRNA targets (browse 221 gene sets)	Gene sets that contain genes sharing putative target sites (seed matches) of human mature miRNA in their 3'-UTRs.	Download GMT Files gene symbols entrez genes ids
TFT: transcription factor targets (browse 615 gene sets)	Gene sets that share upstream *cis*-regulatory motifs which can function as potential transcription factor binding sites. Based on work by Xie et al. 2005	Download GMT Files gene symbols entrez genes ids
C4: computational gene sets (browse 858 gene sets)	Computational gene sets defined by mining large collections of cancer-oriented microarray data. The C4 collection is divided into two sub-collections: CGN and CM. details	Download GMT Files gene symbols entrez genes ids
CGN: cancer gene neighborhoods (browse 427 gene sets)	Gene sets defined by expression neighborhoods centered on 380 cancer-associated genes. This collection is described in Subramanian, Tamayo et al. 2005	Download GMT Files gene symbols entrez genes ids
CM: cancer modules (browse 431 gene sets)	Gene sets defined by Segal et al. 2004. Briefly, the authors compiled gene sets ('modules') from a variety of resources such as KEGG, GO, and others. By mining a large compendium of cancer-related microarray data, they identified 456 such modules as significantly changed in a variety of cancer conditions.	Download GMT Files gene symbols entrez genes ids
C5: GO gene sets (browse 5917 gene sets)	Gene sets that contain genes annotated by the same GO term. The C5 collection is divided into three sub-collections based on GO ontologies: BP, CC, and MF. details	Download GMT Files gene symbols entrez genes ids
BP: GO biological process (browse 4436 gene sets)	Gene sets derived from the GO Biological Process Ontology.	Download GMT Files gene symbols entrez genes ids
CC: GO cellular component (browse 580 gene sets)	Gene sets derived from the GO Cellular Component Ontology.	Download GMT Files gene symbols entrez genes ids
MF: GO molecular function (browse 901 gene sets)	Gene sets derived from the GO Molecular Function Ontology.	Download GMT Files gene symbols entrez genes ids
C6: oncogenic signatures (browse 189 gene sets)	Gene sets that represent signatures of cellular pathways which are often dis-regulated in cancer. The majority of signatures were generated directly from microarray data from NCBI GEO or from internal unpublished profiling experiments involving perturbation of known cancer genes. details	Download GMT Files gene symbols entrez genes ids
C7: immunologic signatures (browse 4872 gene sets)	Gene sets that represent cell states and perturbations within the immune system. The signatures were generated by manual curation of published studies in human and mouse immunology. details	Download GMT Files gene symbols entrez genes ids

But pathways in gene set databases are gene-sets

A pathway



A gene set



1.6. Automatic reconstruction of pathways

Final remark: Automatic reconstruction of pathways

Pathway databases follow two main strategies: Either a curator team, such as in KEGG or Reactome, or open to public submission, such as in Wikipathways.

However, there are huge amounts of pathway information in the scientific literature that would take many years to human beings to process it. Therefore, we need **text mining** methodologies to automatically extract pathway knowledge from the literature.

Final remark: Automatic reconstruction of pathways

One example of this is **MELODI**, a text mining tool that extracts mechanisms of disease based on subject-predicate-object triples from **SemMedDB** (Semantic Medline Database).

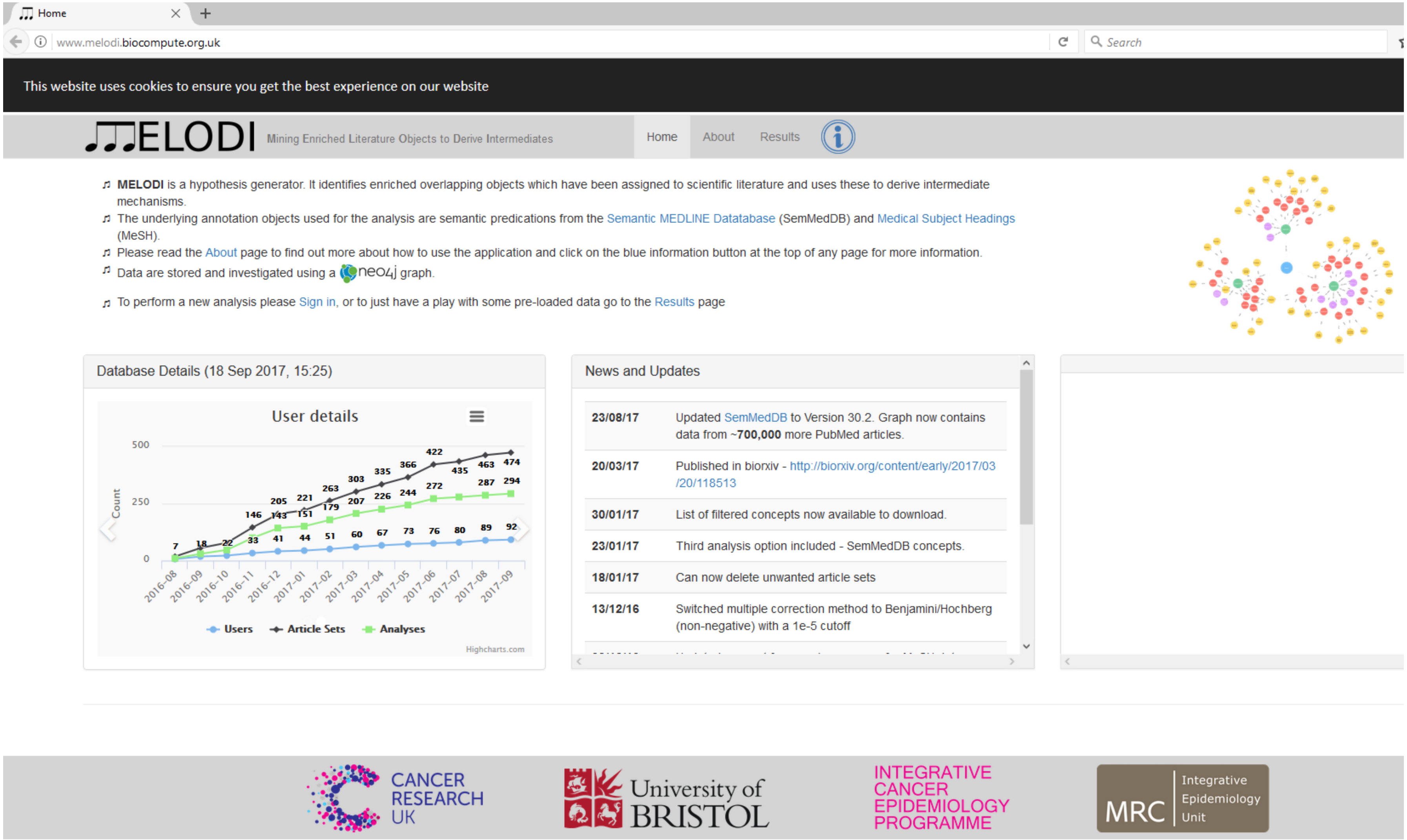
For example, the sentence “*We used hemofiltration to treat a patient with digoxin overdose that was complicated by refractory hyperkalemia*” produces the following four triples:

- Hemofiltration-TREATS-Patients
- Digoxin overdose-PROCESS_OF-Patients
- Hyperkalemia-COMPLICATES-Digoxin overdose
- INFERENCE: Hemofiltration-TREATS-Digoxin overdose

Final remark: Automatic reconstruction of pathways

Building a database of triples for all PubMed, we can let computers link information from different papers and reconstruct the pathway for us!

Final remark: Automatic reconstruction of pathways



Home
www.melodi.biocompute.org.uk
Search

This website uses cookies to ensure your get the best experience on our website

MELODI Mining Enriched Literature Objects to Derive Intermediates
Home
About
Results

- **MELODI** is a hypothesis generator. It identifies enriched overlapping objects which have been assigned to scientific literature and uses these to derive intermediate mechanisms.
- The underlying annotation objects used for the analysis are semantic predications from the Semantic MEDLINE Datatabase (SemMedDB) and Medical Subject Headings (MeSH).
- Please read the About page to find out more about how to use the application and click on the blue information button at the top of any page for more information.
- Data are stored and investigated using a neo4j graph.
- To perform a new analysis please Sign in, or to just have a play with some pre-loaded data go to the Results page

Database Details (18 Sep 2017, 15:25)

News and Updates

23/08/17	Updated SemMedDB to Version 30.2. Graph now contains data from ~**700,000** more PubMed articles.
20/03/17	Published in biorxiv - http://biorxiv.org/content/early/2017/03/20/118513
30/01/17	List of filtered concepts now available to download.
23/01/17	Third analysis option included - SemMedDB concepts.
18/01/17	Can now delete unwanted article sets
13/12/16	Switched multiple correction method to Benjamini/Hochberg (non-negative) with a 1e-5 cutoff

CANCER RESEARCH UK
University of BRISTOL
INTEGRATIVE CANCER EPIDEMIOLOGY PROGRAMME
MRC Integrative Epidemiology Unit

What have we learned today?

What are biological pathways
Where and how to find biological pathways
Pathway database formats and identifiers
How to use the Gene Ontology
What are the main Gene Set databases

Mora Lab

FUNCTIONAL BIOINFORMATICS