Bioinformatic Pathway and Ontology Analysis

David Crossman, Ph.D.
UAB Heflin Center for Genomic Science

Immersion Course
Friday, April 5, 2013

University of Alabama at Birmingham
Heflin Center for Genomic Science

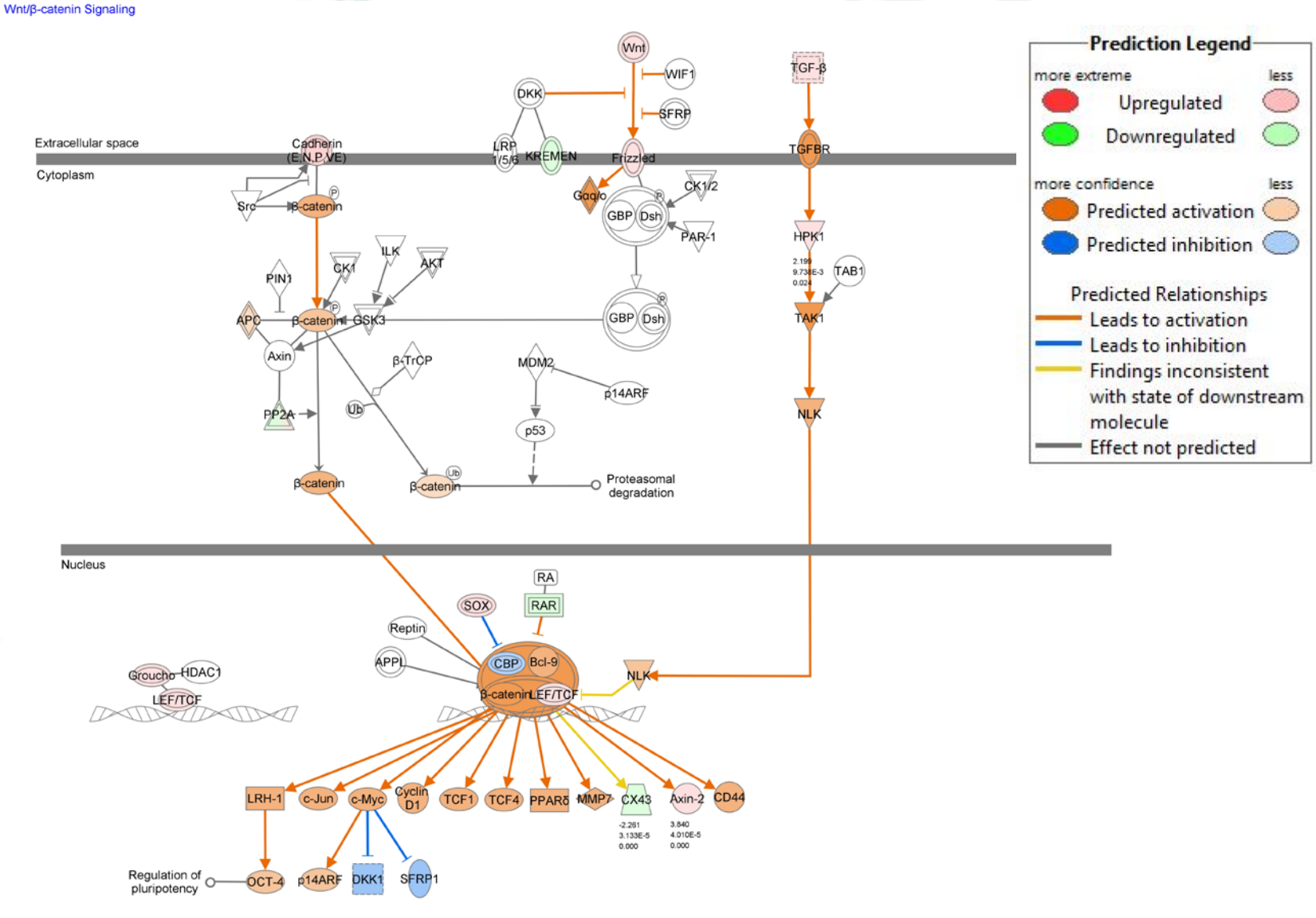
What to do with your list of genes

- Apply a Systems Biology approach to data mine and analyze your data
- Tools and databases available (some free, others $$) to define the underlying biology behind different –omics data
- These tools and databases will identify and prioritize the most relevant pathways, networks and cellular processes affected by your dataset.

Pathways & Ontology Analysis Tools

Tool	Link	Price
Reactome	http://www.reactome.org	Free
IPA	http://ingenuity.com/	$$$
GeneGo Metacore	http://www.genego.com/	$$$
Cytoscape	http://www.cytoscape.org/	Free
GenMAPP	http://www.genmapp.org/	Free
InterMine	http://intermine.org/	Free
KEGG	http://www.genome.jp/kegg/	Free
GO	http://www.geneontology.org/	Free
Panther	http://www.pantherdb.org/	Free
DAVID	http://david.abcc.ncifcrf.gov/	Free
And many, many more!!		

Network Example



© 2000-2012 Ingenuity Systems, Inc. All rights reserved.

Demo

REACTOME

University of Alabama at Birmingham
Heflin Center for Genomic Science

Reactome

- Open-source, open access.
- Manually curated.
- Peer-reviewed pathway database (pathway annotations are authored by “expert” biologists).
- Some of the tools they have:
  - Browse pathways
  - Map IDs to pathways
  - Overrepresentation analysis
  - Compare species
  - Analyze expression data

University of Alabama at Birmingham Heflin Center for Genomic Science

Quick links to tools most commonly used.

Manual & tutorials

Other useful tools

REACTOME

Home | About | Content | Documentation | Tools | Download | Contact Us | Outreach

Search examples...

- Browse Pathways
- Map IDs to Pathways
- Compare Species
- Analyze Expression Data

If you would prefer to use our old website, click here.

Download

The following links allow you to download Reactome data in various formats:

- BioPax
- SBML
- Textbook
- Other formats

Try this

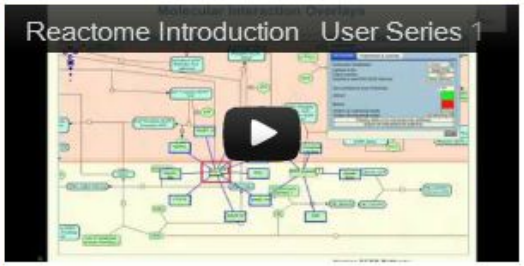
Have you got a set of genes or proteins, where you would like to understand the biological context better? With Reactome, you can find out which of your genes or proteins are overrepresented in which pathways.

Try it out!

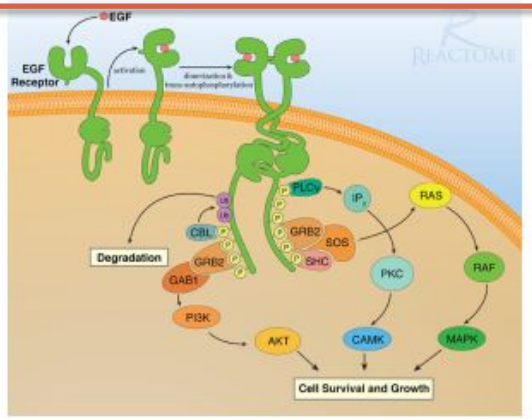
About Reactome

REACTOME is an open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists, in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. These include NCBI Entrez Gene, Ensembl and UniProt databases, the UCSC and HapMap Genome Browsers, the KEGG Compound and ChEBI small molecule databases, PubMed, and Gene Ontology. ... [more]

Tutorial



Featured pathway: Signaling by EGFR



Click image to see pathway

News and Notes

- **Jun 29, 2012: Reactome will be attending COMBINE 2012**

The Computational Modeling in Biology Network (COMBINE) is an initiative to coordinate the development of the various community standards ... [more]

- **Jun 29, 2012: Reactome will be visiting Korea and Japan**

Reactome outreach coordinator, Robin Haw, will give a talk at the Current and Future in Pathway Research International Workshop at the Korea ... [more]

- **Jun 26, 2012: Version 41 Released**

New in Version 41 is the topic Cellular responses to stress, including Regulation of Hypoxia-inducible ... [more]

- **Mar 22, 2012: Version 40 Released**

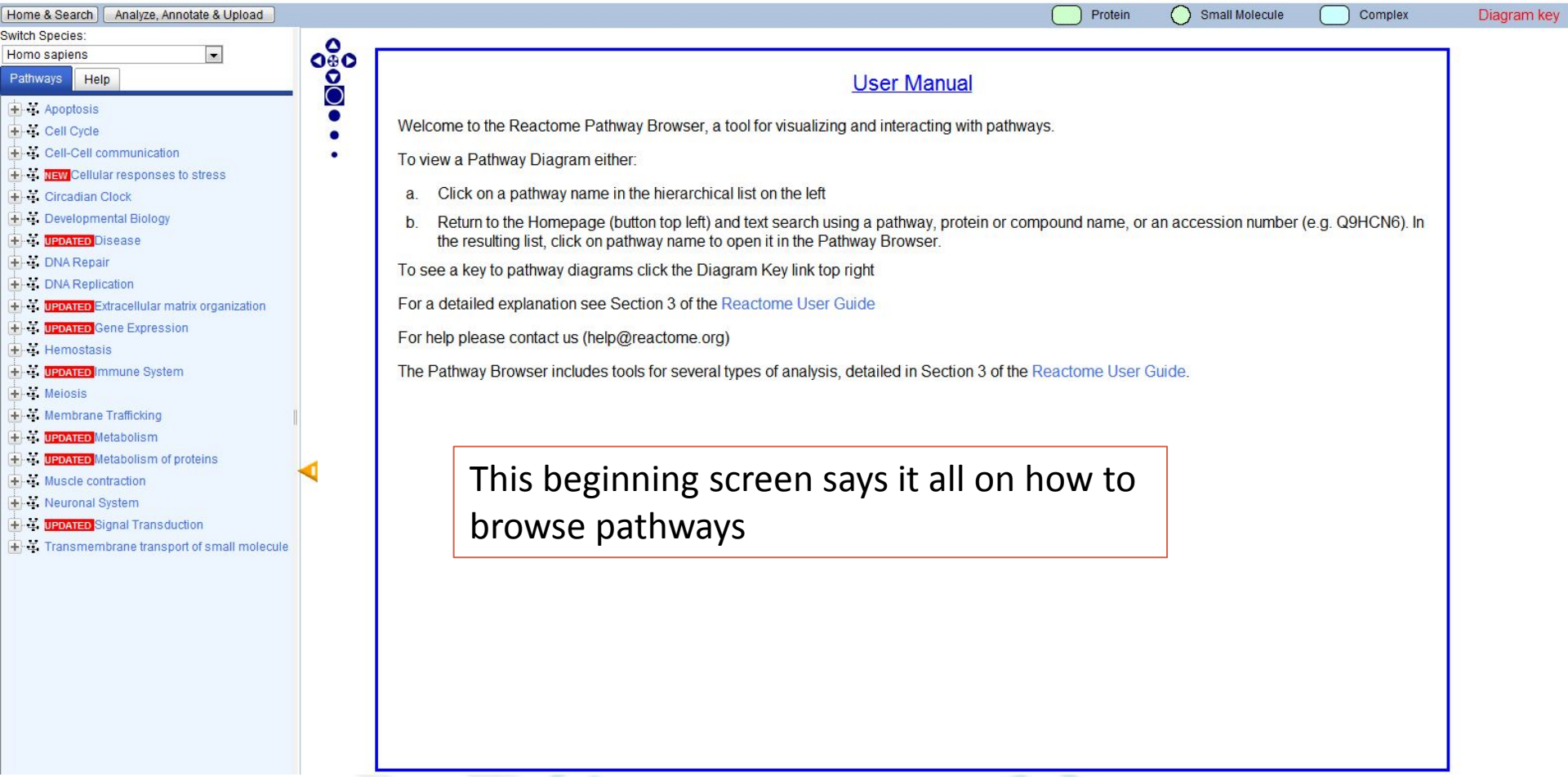
The pathway Extracellular matrix organization is new and includes Activation ... [more]

- **Mar 16, 2012: Reactome will be participating in Google Summer of Code 2012**

The Genome Informatics group, organizing the joint efforts of Galaxy, GBrowse, ... [more]

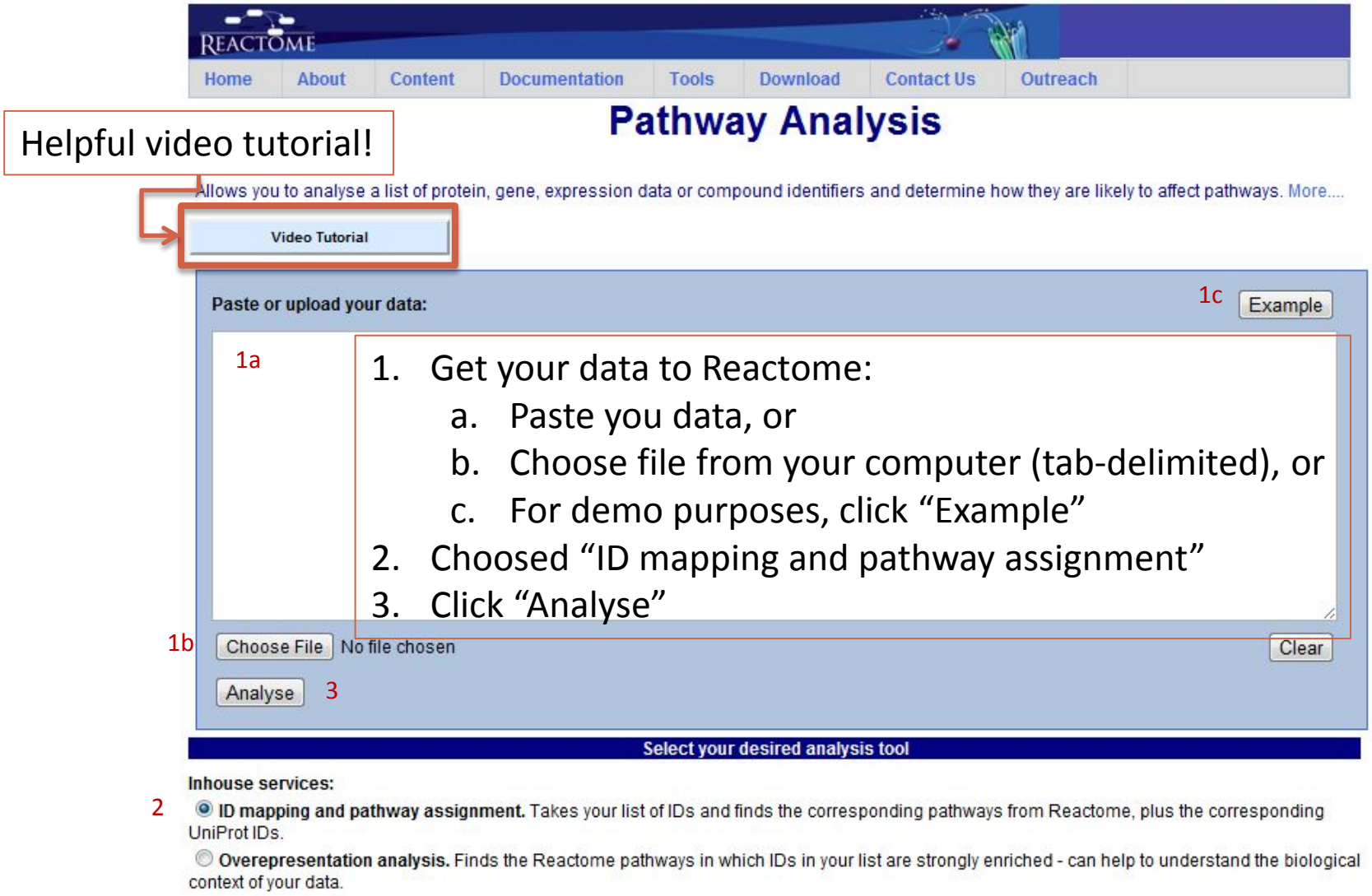
To see older news, click here.

Browse Pathways

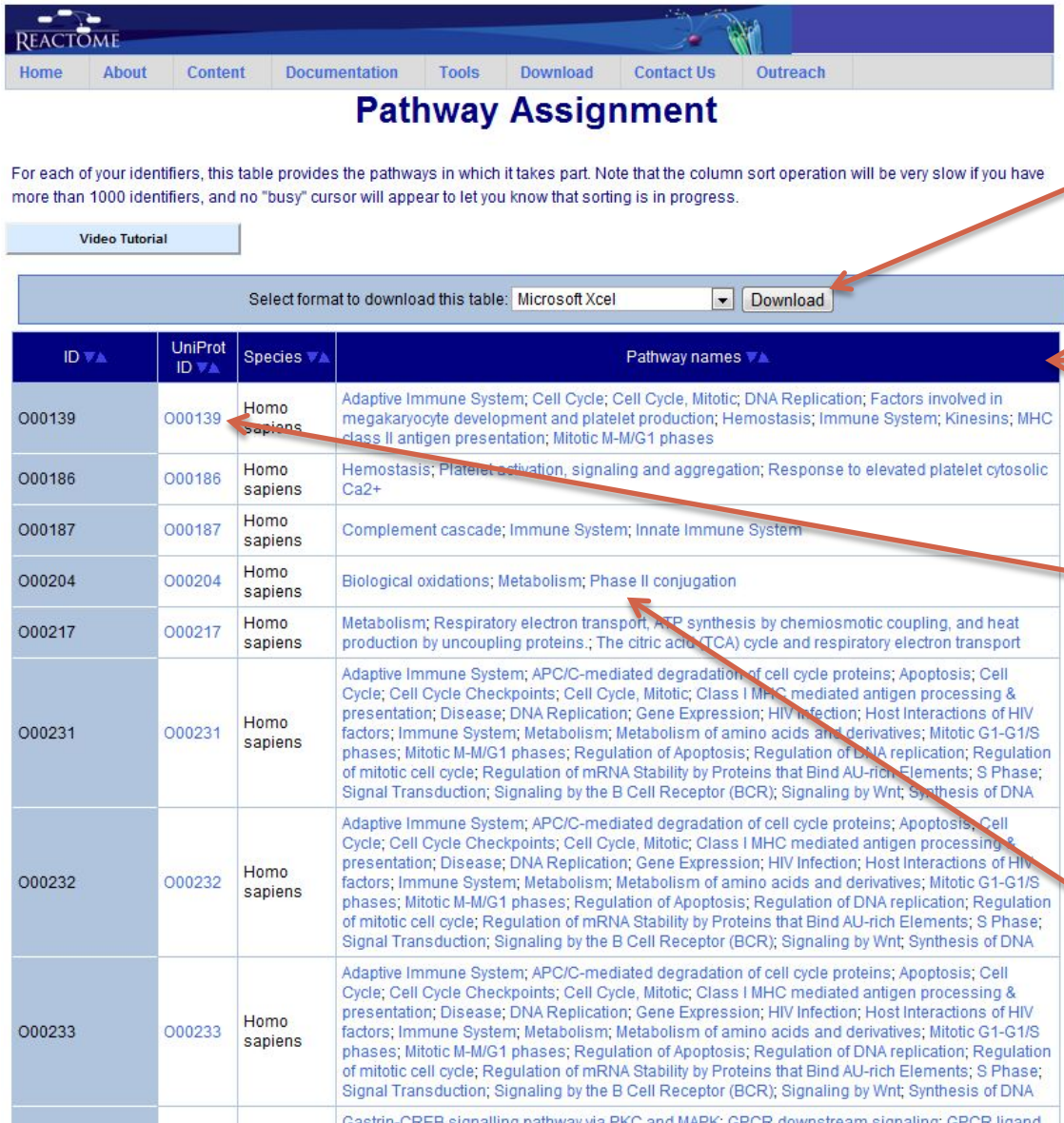


This beginning screen says it all on how to browse pathways

Map IDs to Pathways
(ID mapping and pathway assignment)



ID mapping & pathway assignment



REACTOME

Home | About | Content | Documentation | Tools | Download | Contact Us | Outreach

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.

Video Tutorial

Select format to download this table: Microsoft Xcel Download

ID	UniProt ID	Species	Pathway names
O00139	O00139	Homo sapiens	Adaptive Immune System; Cell Cycle; Cell Cycle, Mitotic; DNA Replication; Factors involved in megakaryocyte development and platelet production; Hemostasis; Immune System; Kinesins; MHC class II antigen presentation; Mitotic M-M/G1 phases
O00186	O00186	Homo sapiens	Hemostasis; Platelet activation, signaling and aggregation; Response to elevated platelet cytosolic Ca2+
O00187	O00187	Homo sapiens	Complement cascade; Immune System; Innate Immune System
O00204	O00204	Homo sapiens	Biological oxidations; Metabolism; Phase II conjugation
O00217	O00217	Homo sapiens	Metabolism; Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.; The citric acid (TCA) cycle and respiratory electron transport
O00231	O00231	Homo sapiens	Adaptive Immune System; APC/C-mediated degradation of cell cycle proteins; Apoptosis; Cell Cycle; Cell Cycle Checkpoints; Cell Cycle, Mitotic; Class I MHC mediated antigen processing & presentation; Disease; DNA Replication; Gene Expression; HIV Infection; Host Interactions of HIV factors; Immune System; Metabolism; Metabolism of amino acids and derivatives; Mitotic G1-G1/S phases; Mitotic M-M/G1 phases; Regulation of Apoptosis; Regulation of DNA replication; Regulation of mitotic cell cycle; Regulation of mRNA Stability by Proteins that Bind AU-rich Elements; S Phase; Signal Transduction; Signaling by the B Cell Receptor (BCR); Signaling by Wnt; Synthesis of DNA
O00232	O00232	Homo sapiens	Adaptive Immune System; APC/C-mediated degradation of cell cycle proteins; Apoptosis; Cell Cycle; Cell Cycle Checkpoints; Cell Cycle, Mitotic; Class I MHC mediated antigen processing & presentation; Disease; DNA Replication; Gene Expression; HIV Infection; Host Interactions of HIV factors; Immune System; Metabolism; Metabolism of amino acids and derivatives; Mitotic G1-G1/S phases; Mitotic M-M/G1 phases; Regulation of Apoptosis; Regulation of DNA replication; Regulation of mitotic cell cycle; Regulation of mRNA Stability by Proteins that Bind AU-rich Elements; S Phase; Signal Transduction; Signaling by the B Cell Receptor (BCR); Signaling by Wnt; Synthesis of DNA
O00233	O00233	Homo sapiens	Adaptive Immune System; APC/C-mediated degradation of cell cycle proteins; Apoptosis; Cell Cycle; Cell Cycle Checkpoints; Cell Cycle, Mitotic; Class I MHC mediated antigen processing & presentation; Disease; DNA Replication; Gene Expression; HIV Infection; Host Interactions of HIV factors; Immune System; Metabolism; Metabolism of amino acids and derivatives; Mitotic G1-G1/S phases; Mitotic M-M/G1 phases; Regulation of Apoptosis; Regulation of DNA replication; Regulation of mitotic cell cycle; Regulation of mRNA Stability by Proteins that Bind AU-rich Elements; S Phase; Signal Transduction; Signaling by the B Cell Receptor (BCR); Signaling by Wnt; Synthesis of DNA

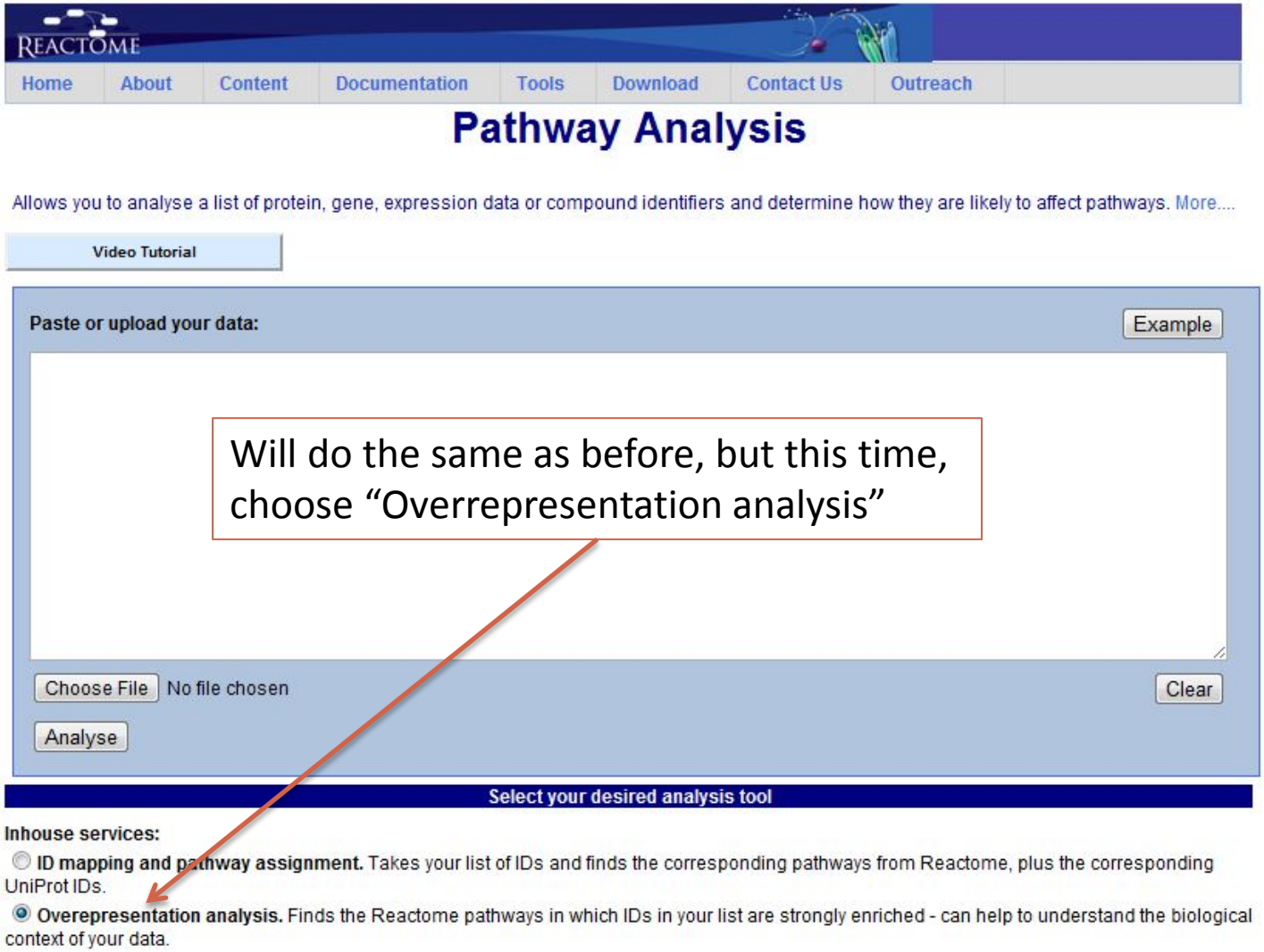
Table can be downloaded in various formats

Columns can be sorted

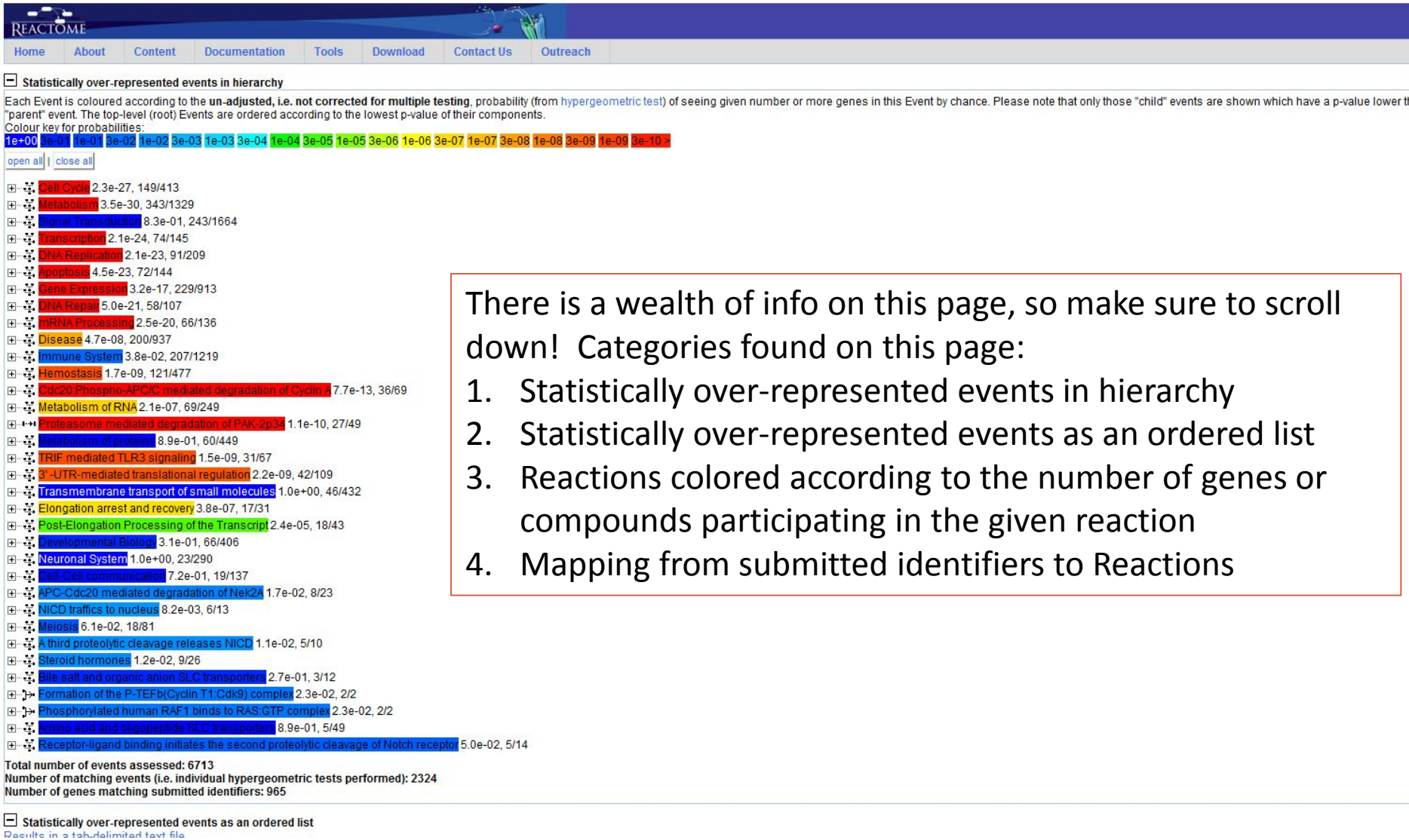
Clicking the UniProt ID # will take you to the info page about that particular molecule

Clicking on the Pathway names links will take you to the pathway map

Map IDs to Pathways (Overrepresentation analysis)



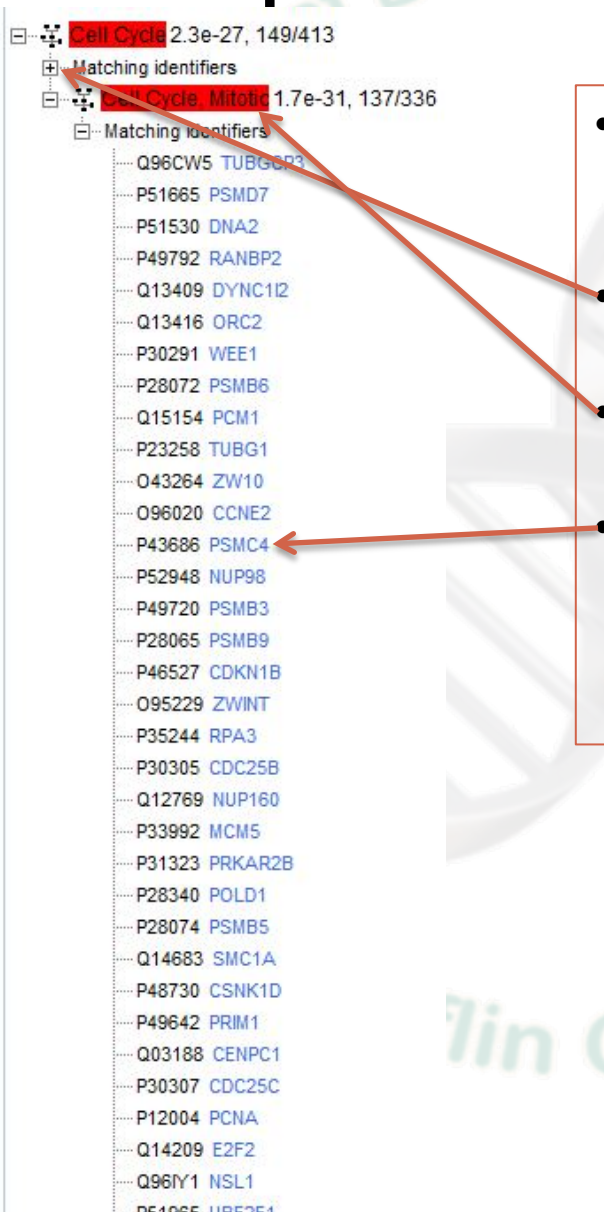
Overrepresentation analysis



There is a wealth of info on this page, so make sure to scroll down! Categories found on this page:

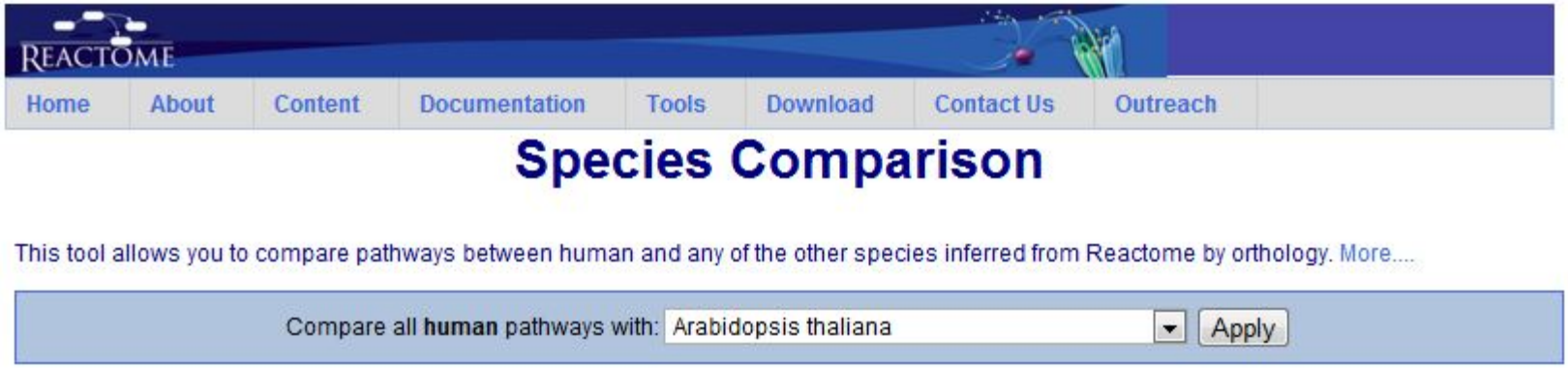
1. Statistically over-represented events in hierarchy
2. Statistically over-represented events as an ordered list
3. Reactions colored according to the number of genes or compounds participating in the given reaction
4. Mapping from submitted identifiers to Reactions

Overrepresentation analysis



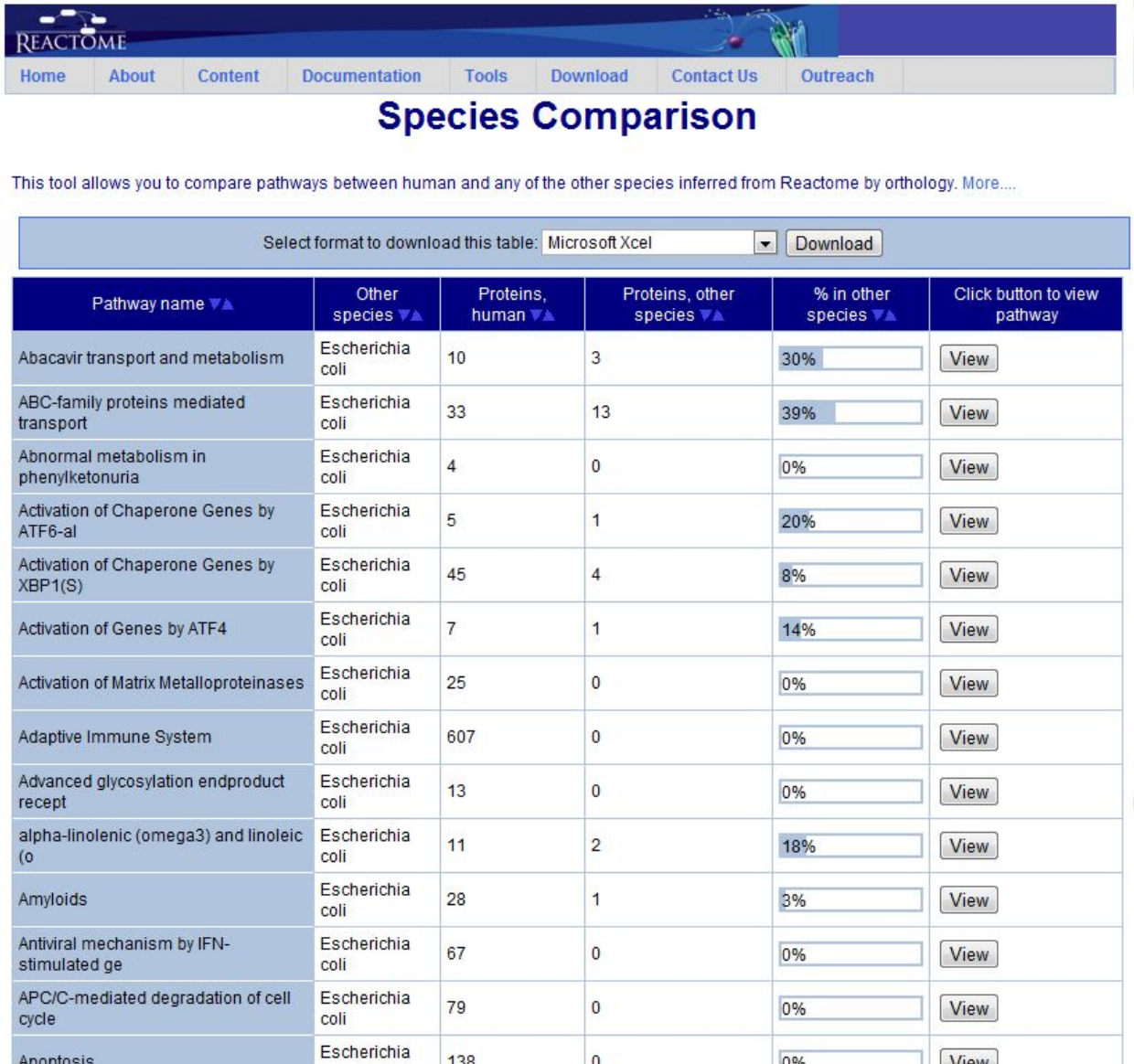
- Here, we are focusing in on the “Statistically over-represented events in hierarchy” category.
- Clicking on the “+” sign will expand the event (“-” will close it)
- Clicking on the colored words will open the pathway map
- Clicking on any gene name link will take you to an info page about that gene

Compare species



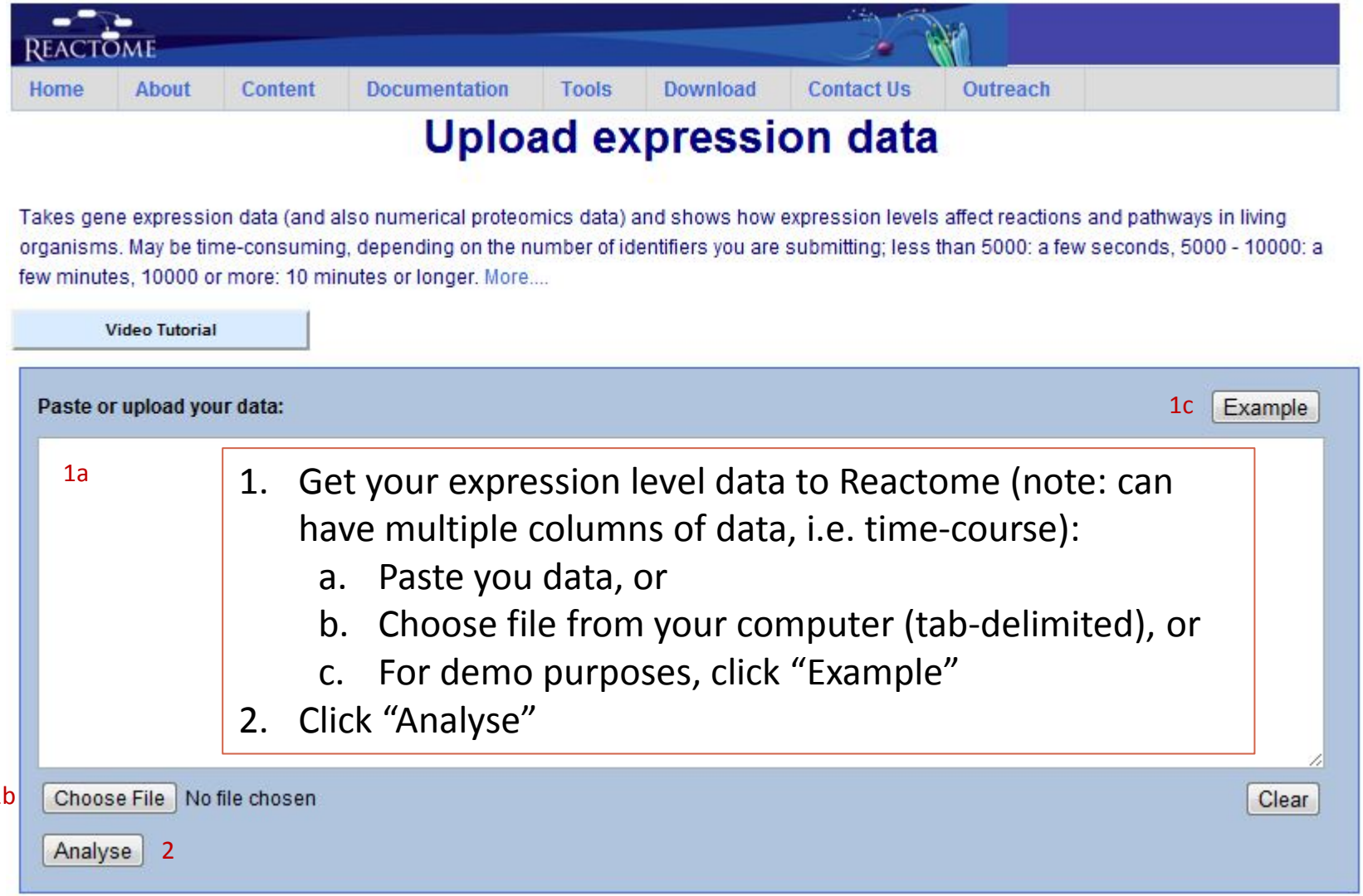
This tool will allow you to compare human pathways to any other species pathways they have in their database

Compare species

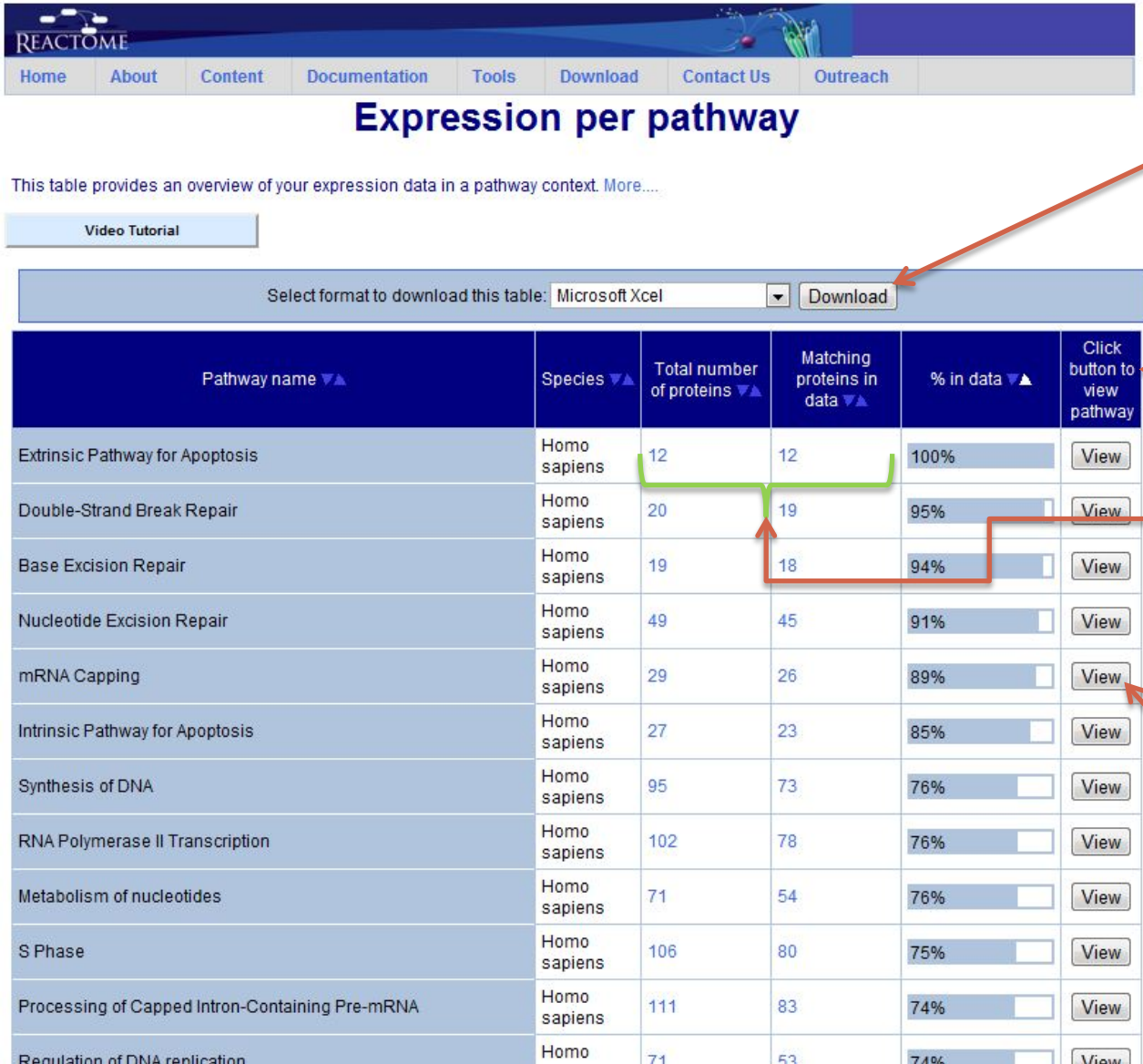


Here I chose, *E. coli*, and as before, columns can be sorted, the “View” button will open the pathway map, and the table can be downloaded

Analyze expression data



Analyze expression data



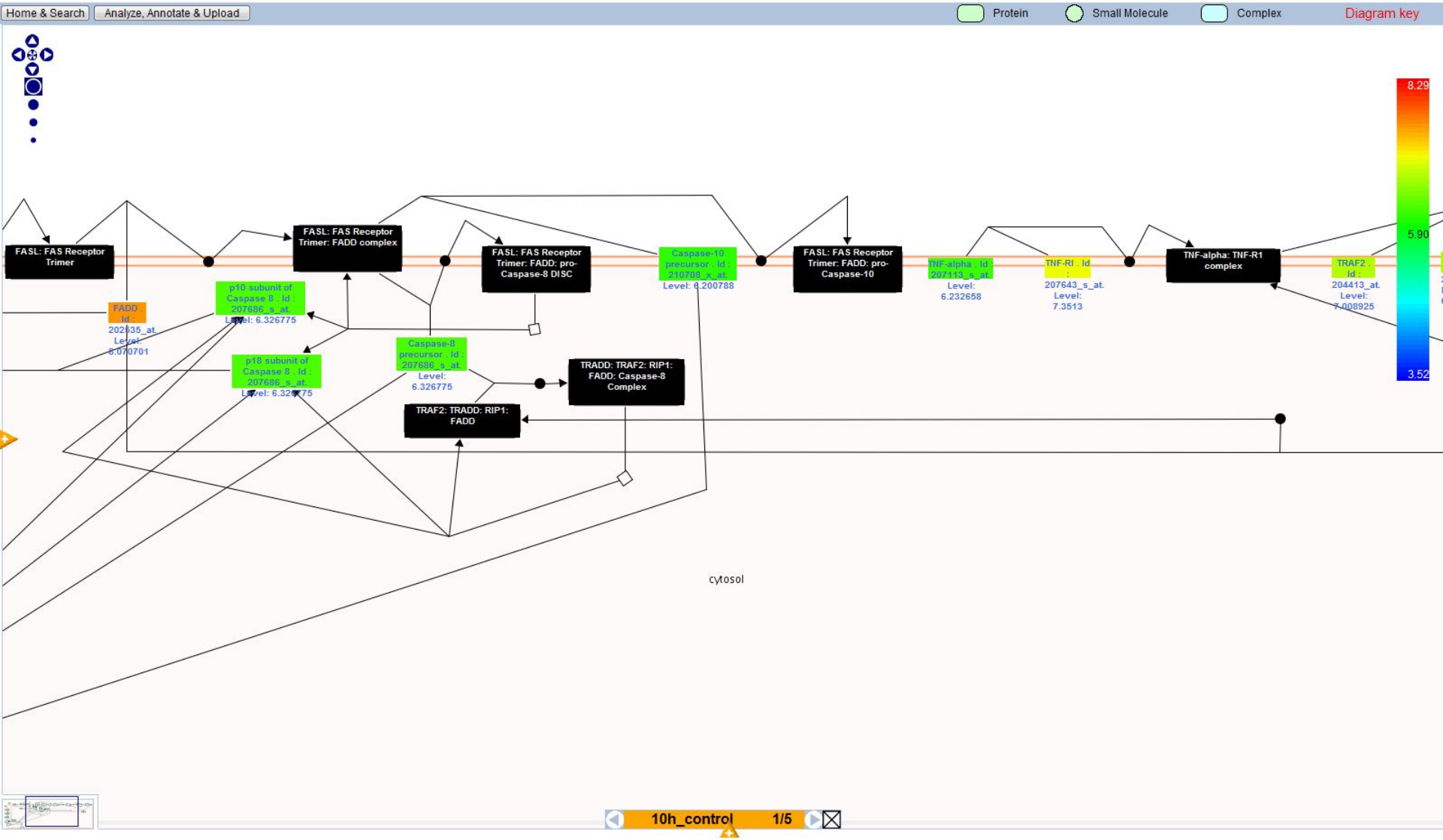
Pathway name	Species	Total number of proteins	Matching proteins in data	% in data	Click button to view pathway
Extrinsic Pathway for Apoptosis	Homo sapiens	12	12	100%	View
Double-Strand Break Repair	Homo sapiens	20	19	95%	View
Base Excision Repair	Homo sapiens	19	18	94%	View
Nucleotide Excision Repair	Homo sapiens	49	45	91%	View
mRNA Capping	Homo sapiens	29	26	89%	View
Intrinsic Pathway for Apoptosis	Homo sapiens	27	23	85%	View
Synthesis of DNA	Homo sapiens	95	73	76%	View
RNA Polymerase II Transcription	Homo sapiens	102	78	76%	View
Metabolism of nucleotides	Homo sapiens	71	54	76%	View
S Phase	Homo sapiens	106	80	75%	View
Processing of Capped Intron-Containing Pre-mRNA	Homo sapiens	111	83	74%	View
Regulation of DNA replication	Homo	71	53	74%	View

Table can be downloaded in various formats

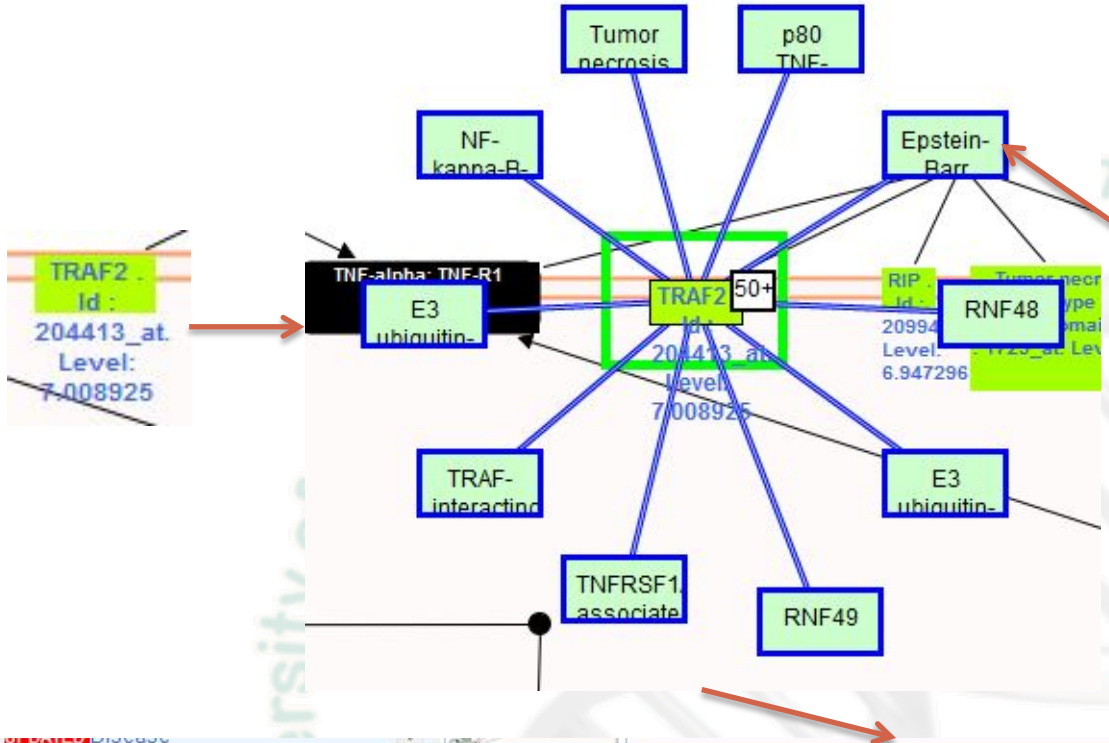
Columns can be sorted

Clicking the number link will show list of molecules involved

Clicking on the “View” button will take you to the pathway map

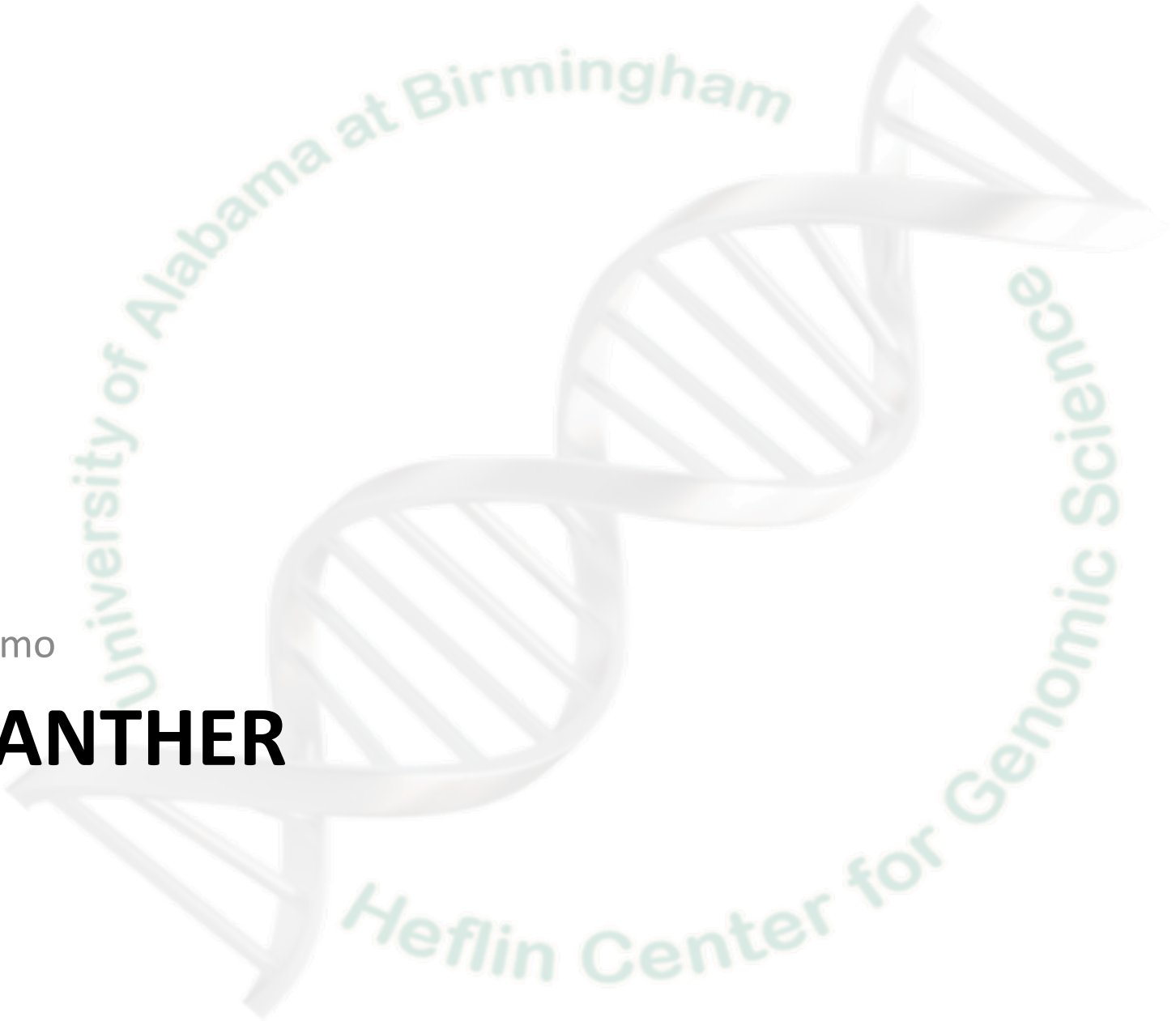


- Here I chose the Extrinsic Pathway for Apoptosis
- Molecules colored are those from the dataset (values are listed below the molecule name)
- Click on Diagram key (top right hand corner) to reveal the legend



- Molecules are interactive
- Click on a molecule (a green box will highlight it)
- Right-click and choose “Display interactors”
- Click on yellow triangle with a “+” sign at bottom to reveal detailed info about the molecule

Name	TRAF2 TNF receptor associated factor 2 Tumor necrosis factor type 2 receptor associated protein 3
Stable identifier	REACT_4499.2
Links to corresponding entries in other databases	BioGPS Gene:7186 CTD Gene:7186 ENSEMBL:ENSG00000127191 GeneCards:Q12933 KEGG Gene:7186 NCBI Gene:7186 OMIM:601895 Protein Data Bank:1CA4 Protein Data Bank:1CA9 Protein Data Bank:1CZY Protein Data Bank:1CZZ Protein Data Bank:1D00 Protein Data Bank:1D01 Protein Data Bank:1D0A Protein Data Bank:1D0J Protein Data Bank:1F3V Protein Data Bank:1QSC RefSeq:NP_066961 UCSC human:Q12933



Demo

PANTHER

Panther

- Tools and data on the PANTHER site can be used to:
  - Get information about a gene of interest
  - Explore protein families, molecular functions, biological processes, cellular components and pathways
  - Generate lists of genes that belong to a given protein family or subfamily, have a given molecular function or participate in a given biological process or pathway, e.g. generate a candidate gene list for a disease
  - Analyze lists of genes, proteins or transcripts according to categories based on family, molecular function, biological process, cellular component or pathway, e.g. analyze mRNA microarray data

Other great features

LOGIN REGISTER CONTACT US HELP

Home | Browse | Genes and orthologs | Trees and HMMs | Pathways | Ontologies | Tools | Workspace

Quick links

- Whole genome function views
- Gene expression tools
- cSNP tools
- Upload multiple gene IDs
- Community Curation
- My Workspace
- HMM scoring
- Downloads
- Genome statistics
- Site map

Newsletter subscription

Enter your Email:

Subscribe

Quick links to tools most commonly used.

KEYWORD SEARCH

All Go

SEQUENCE SEARCH

Enter a protein sequence:

Sequence query limits: Protein - 50kb

Submit

The PANTHER (**P**rotein **AN**alysis **TH**rough **E**volutionary **R**elationships) Classification System is a unique resource that **classifies genes by their functions**, using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct experimental evidence. Proteins are **classified by expert biologists** according to:

- Gene families and subfamilies, including annotated phylogenetic trees
- Gene Ontology classes: molecular function, biological process, cellular component
- PANTHER Protein Classes
- Pathways, including diagrams

PANTHER is part of the Gene Ontology Reference Genome Project.

PANTHER is supported by a research grant from the National Institute of General Medical Sciences [grant GM081084] and maintained by the Thomas lab at the University of Southern California.

What can I do on the PANTHER site?
Guide to getting started

News

(March 16, 2012)

PANTHER 7.2 is released.

Click for additional info.

Publications

How to cite PANTHER

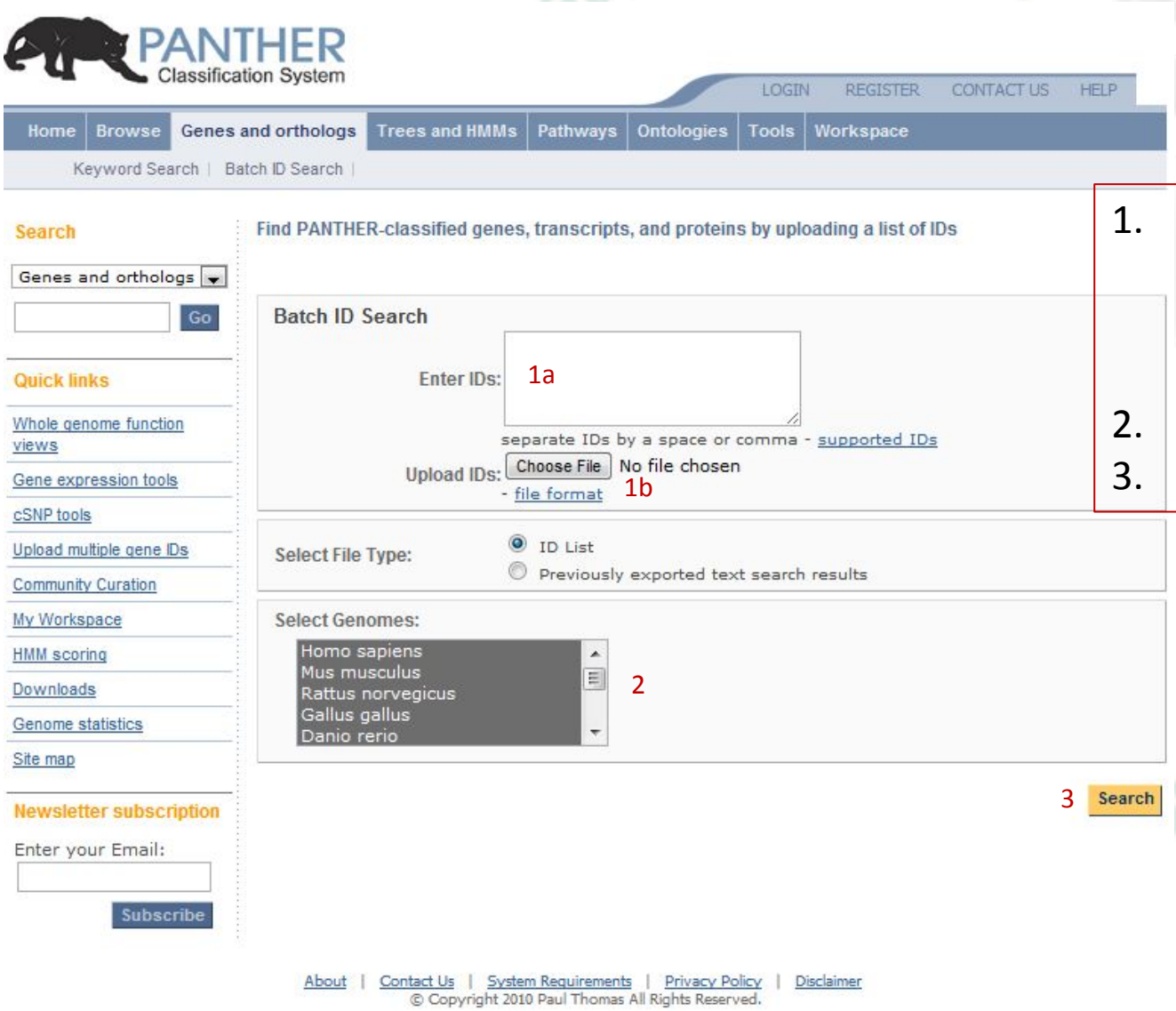
"PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium." Mi, et al.

"Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools." Thomas, et al.

"PANTHER: a library of protein families and subfamilies indexed by function." Thomas, et al.

© Copyright 2010 Paul Thomas All Rights Reserved.

Upload multiple gene IDs



1. Get your data to Panther:
   a. Paste your IDs, or
   b. Choose a file (tab-delimited)
2. Select Genome(s)
3. Click “Search”

- Table listing all the genes in your dataset (scroll to the right to see all the columns
- Table can be downloaded by the “Send list to:” dropdown box
- Click the pie chart icon to view the GO categories in pie charts

PANTHER Classification System

Home | Browse | Genes and orthologs | Trees and HMMs | Pathways | Ontologies | Tools | Workspace

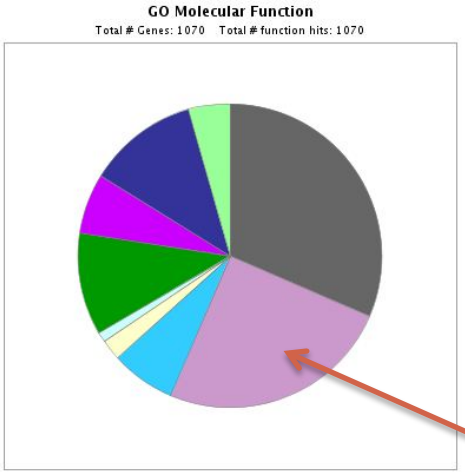
PANTHER GENE LIST Customize Gene list

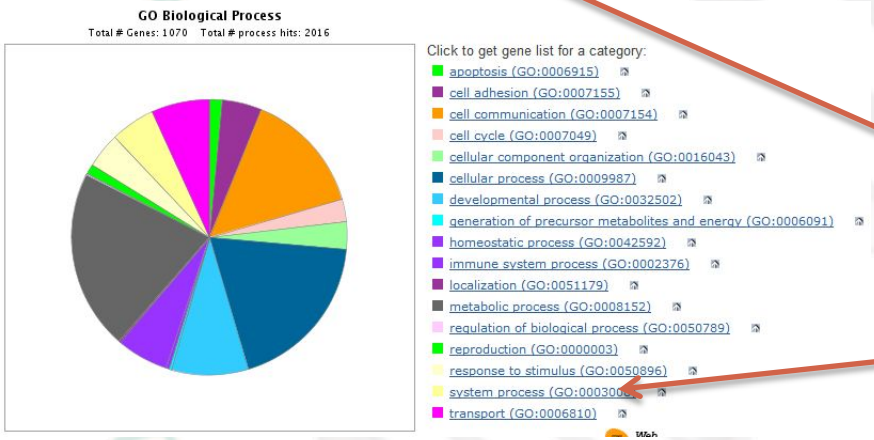
Convert List to: -Select- Send list to: -Select-

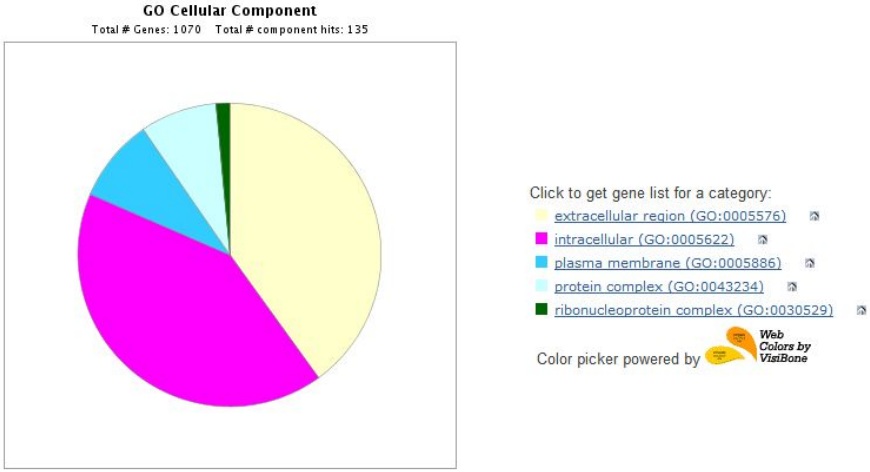
Display: 30 items per page Refine Search

Hits 1-30 of 1070 [page: (1) 2 3 4 5 6 7 8 9 10 >>] Number of mapped ids found 1040 IDs not found (217)

clr all		Gene ID	Mapped IDs	Gene Name Gene Symbol Ortholog	PANTHER Family/Subfamily	GO Molecular Function	GO Biological Proc
	1.	HUMAN|ENSEMBL=ENSG00000115361|UniProtKB=P28330	ACADL	Long-chain specific acyl-CoA dehydrogenase, mitochondrial ACADL ortholog	PREDICTED: SIMILAR TO ISOVALERYL COENZYME A DEHYDROGENASE, PARTIAL (PTHR10909:SF143)	oxidoreductase activity	respiratory electro acyl-CoA metaboli acyl-CoA metaboli
	2.	HUMAN|ENSEMBL=ENSG00000005471|UniProtKB=P21439	ABCB4	Multidrug resistance protein 3 ABCB4 ortholog	SUBFAMILY NOT NAMED (PTHR24221:SF24)	ATPase activity, coupled to transmembrane movement of substances transmembrane transporter activity	extracellular trans
	3.	HUMAN|ENSEMBL=ENSG00000155893|UniProtKB=Q8TE99	ACPL2	Acid phosphatase-like protein 2 ACPL2 ortholog	HISTIDINE ACID PHOSPHATASE-RELATED (PTHR11567:SF12)	-	-
	4.	HUMAN|ENSEMBL=ENSG00000197142|UniProtKB=Q9ULC5	ACSL5	Long-chain-fatty-acid--CoA ligase 5 ACSL5 ortholog	PUTATIVE UNCHARACTERIZED PROTEIN (PTHR24096:SF46)	ligase activity	fatty acid metaboli
	5.	HUMAN|ENSEMBL=ENSG00000160345|UniProtKB=Q5BN46	C9orf116	UPF0691 protein C9orf116 C9orf116 ortholog	SUBFAMILY NOT NAMED (PTHR20899:SF0)	-	-
	6.	HUMAN|ENSEMBL=ENSG00000165118|UniProtKB=Q5T6V5	C9orf64	UPF0553 protein C9orf64 C9orf64 ortholog	SUBFAMILY NOT NAMED (PTHR21314:SF0)	-	-
	7.	HUMAN|ENSEMBL=ENSG00000182985|UniProtKB=Q9BY67	CADM1	Cell adhesion molecule 1	NECTIN-RELATED (PTHR23277:SF8)	receptor activity	-







- Pie charts of the 3 GO categories:
  - Molecular Function
  - Biological Process
  - Cellular Component
- Each “wedge” can be clicked on to drill further down into the cateogry
- Clicking on any category name link will list a table of molecules from your dataset found in that particular category