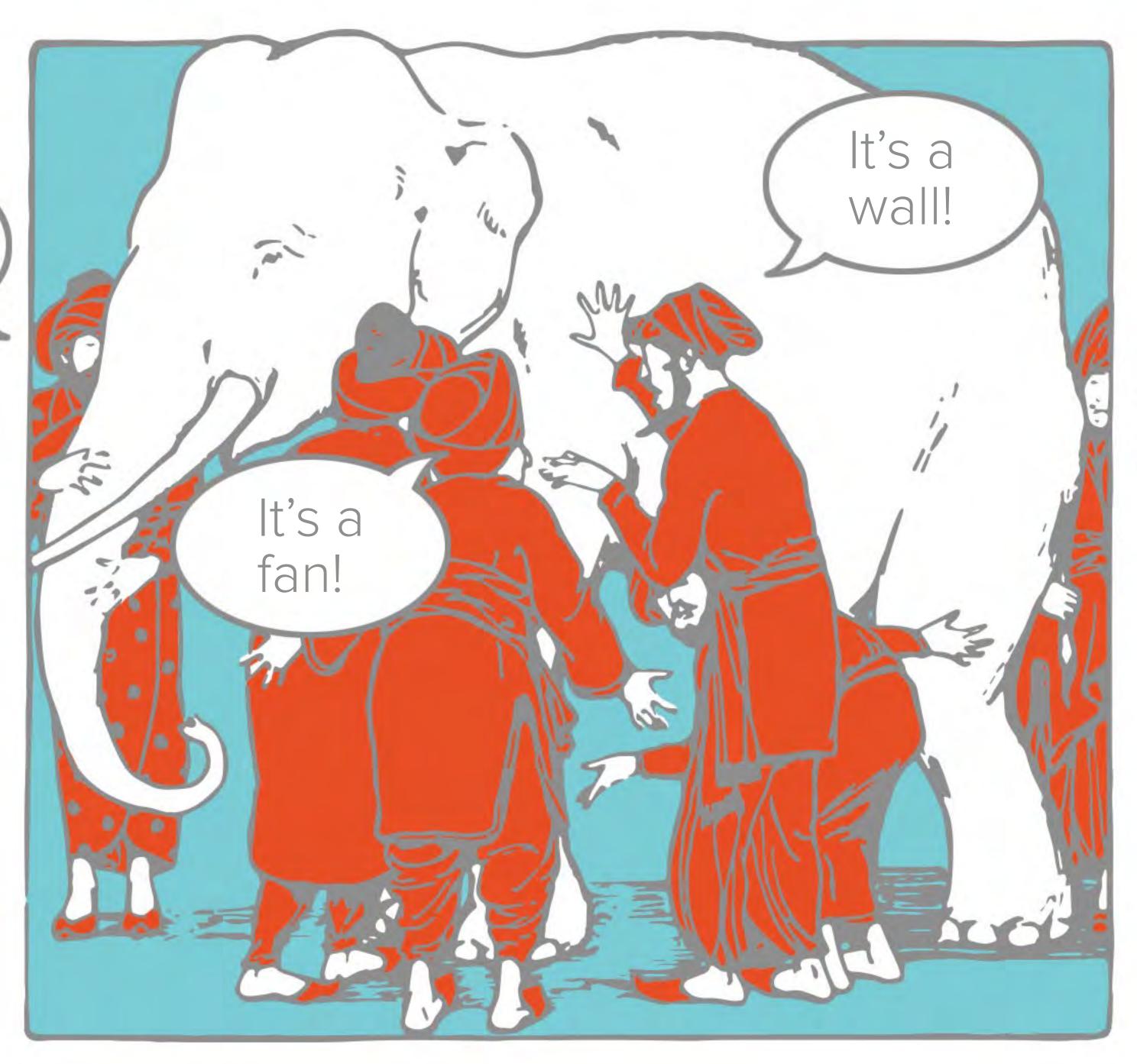
# Introduction to Systems Biology

- Definitions and fundamental concepts
- Tools for Systems Biology analysis overview
- Pathway Analysis: overview
- Gene Ontology Analysis: overview
- Public on-line resources hands-on

It's a snake!

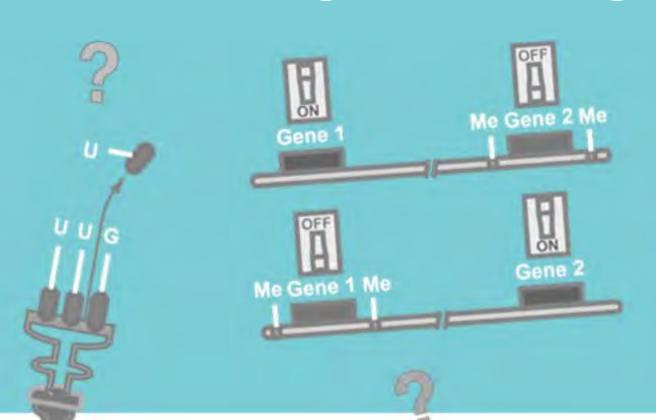


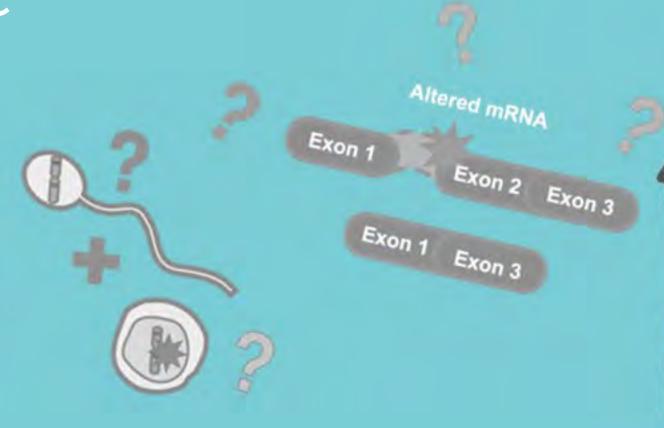
# Blind men & the Elephant

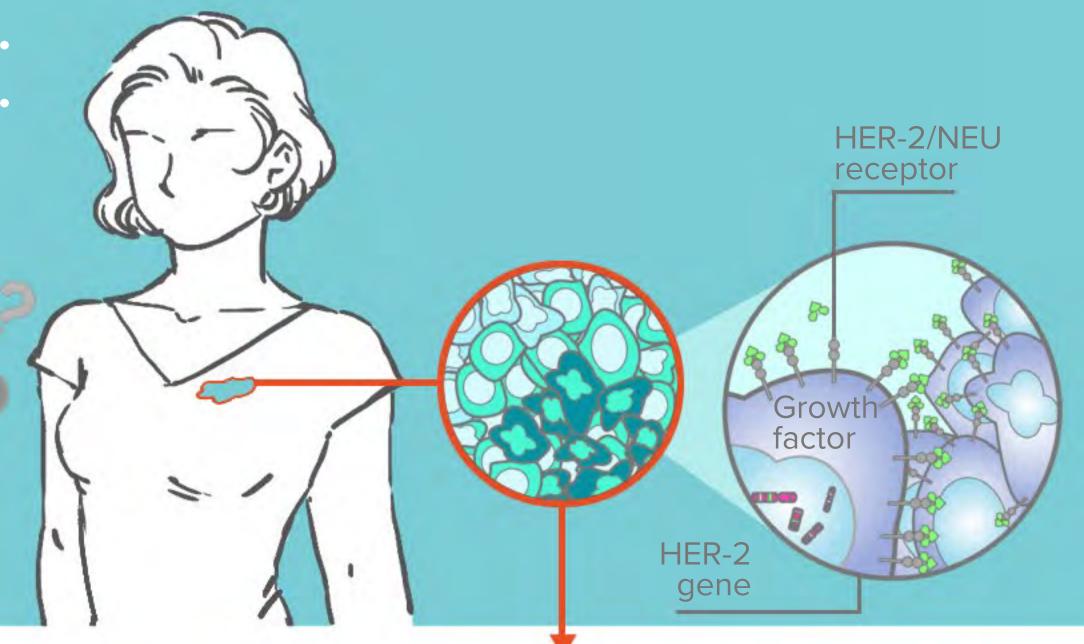
It's a rope!

Personalized medicine:

A daunting challenge







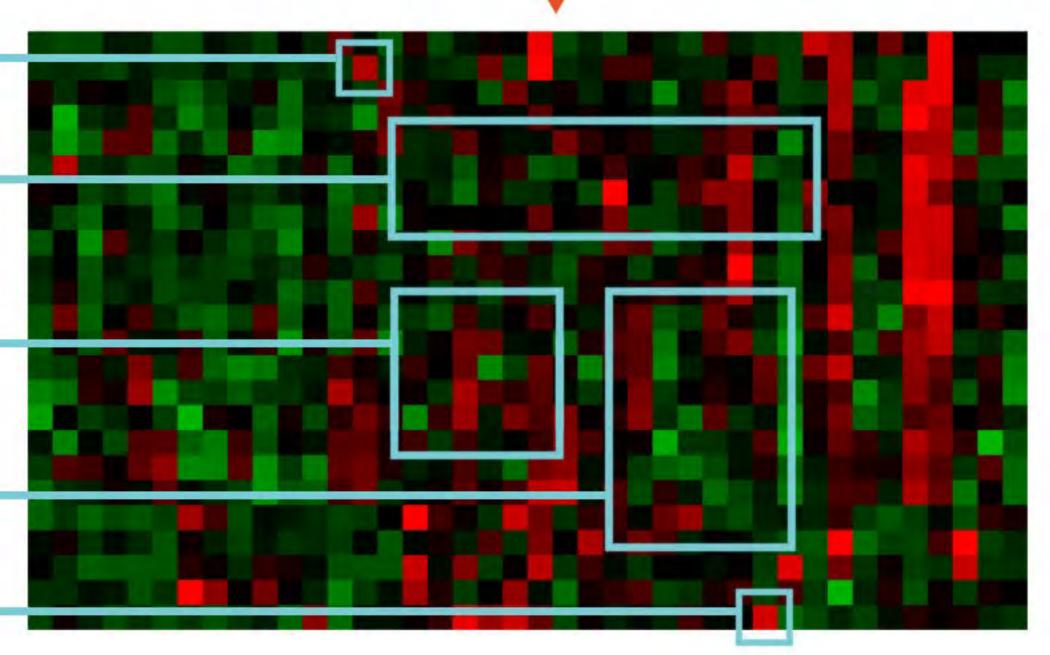
Growth receptor present

High risk of relapse

Not good candidate for conventional therapy

Mutated susceptibility genes

Vascular markers present



## Systems Biology – Definitions

To understand complex biological systems requires the integration of experimental and computational research — in other words a systems biology approach. (Kitano, 2002)

Systems biology studies biological systems by systematically perturbing them (biologically, genetically, or chemically); monitoring the gene, protein, and informational pathway responses; integrating these data; and ultimately, formulating mathematical models that describe the structure of the system and its response to individual perturbations. (Ideker et al, 2001)

[...]the objective of systems biology [can be] defined as the understanding of network behavior, and in particular their dynamic aspects, which requires the utilization of mathematical modeling tightly linked to experiment. (Cassman, 2005)

## Systems Biology — Definitions

By discovering how function arises in dynamic interactions, systems biology addresses the missing links between molecules and physiology. Top-down systems biology identifies molecular interaction networks on the basis of correlated molecular behavior observed in genome-wide "omics" studies. Bottom-up systems biology examines the mechanisms through which functional properties arise in the interactions of known components. (Bruggeman and Westerhoff, 2007)

### Systems Biology – Definitions

# Systems biology is the study of systems of biological components, which may be molecules, cells, organisms or entire species.

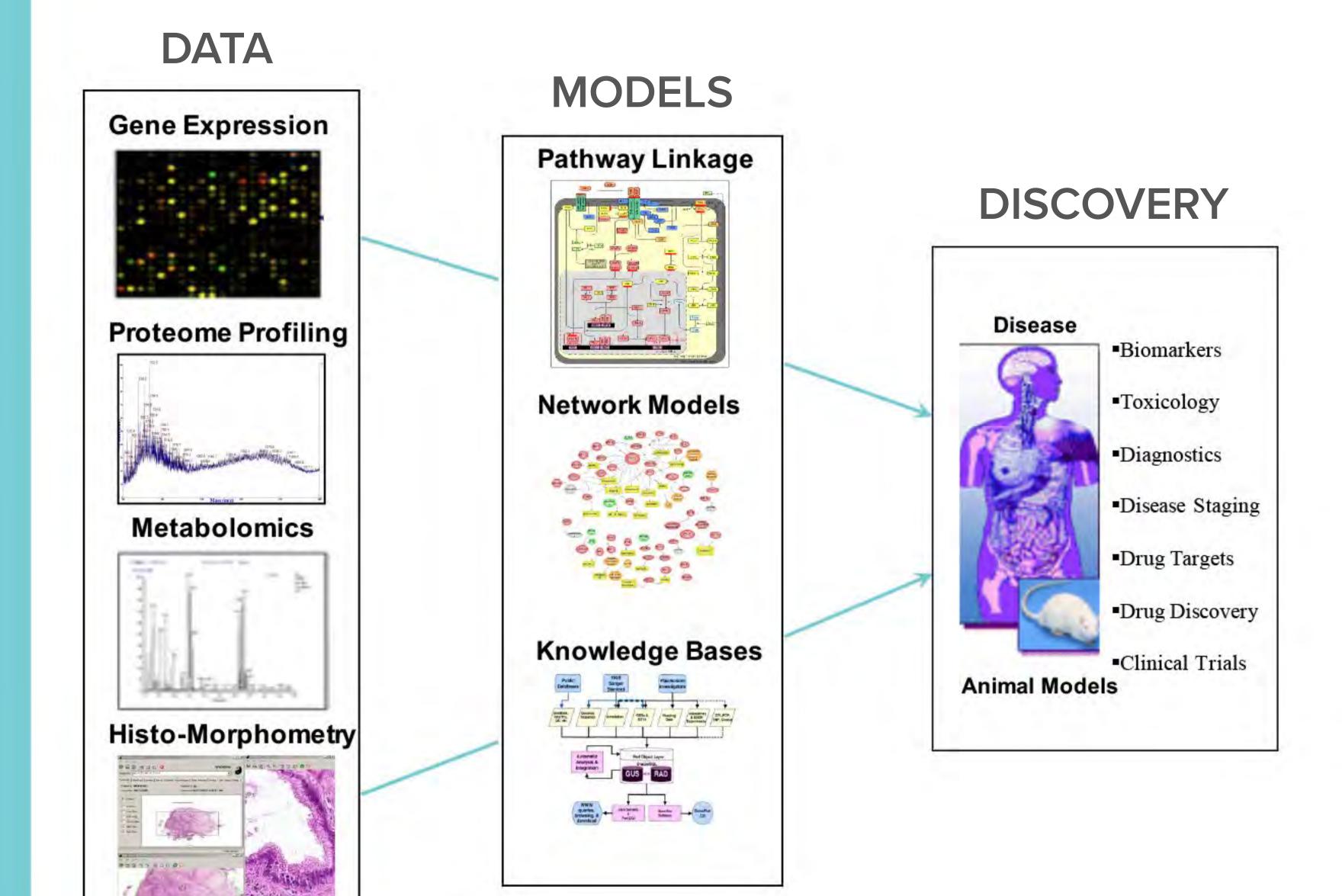
Living systems are dynamic and complex, and their behavior may be hard to predict from the properties of individual parts.

To study them, we use quantitative measurements of the behavior of groups of interacting components, systematic measurement technologies such as genomics, bioinformatics and proteomics, and mathematical and computational models to describe and predict dynamical behavior.

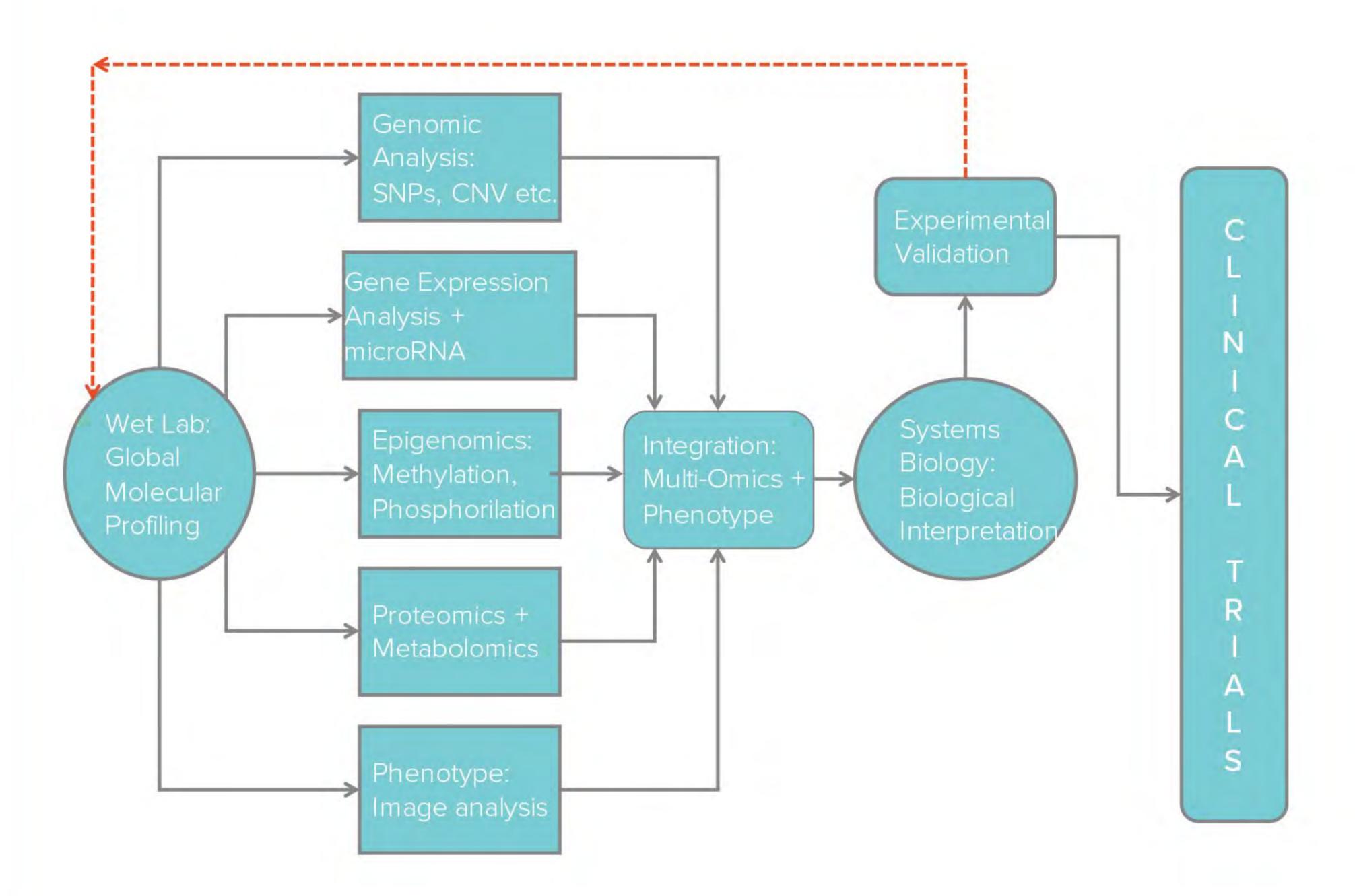
# Systems problems are emerging as central to all areas of biology and medicine.

- Dep. Systems Biology, Harvard University

Systems Biology /
Integrative Bioinformatics



How does it all fit together?



## Systems Biology Methods/ Analysis tools

#### Pathway Analysis

Biological Pathways/ Signaling Pathways - overview

Bioinformatics Tools and Resources for Pathway Analysis

Pathway Studio (Ariadne/Elsevier)

GeneGo (Thomson Scientific/Reuters)

Ingenuity Pathway Analysis IPA (Ingenuity Systems)

Applications: micro-array and other omics- data analysis and text mining

Gene Ontology – GO Enrichment Analysis

Biological Interactions Networks

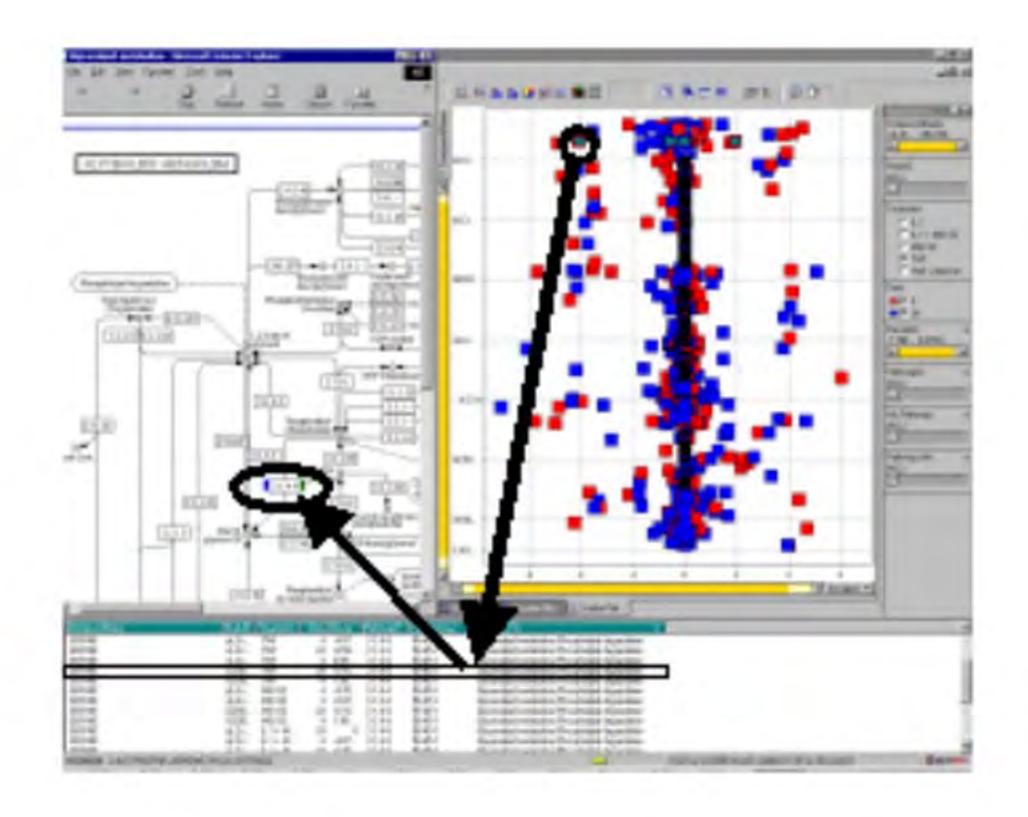
From Pathways and Networks to Biological Interpretation

# Pathway Resources Central

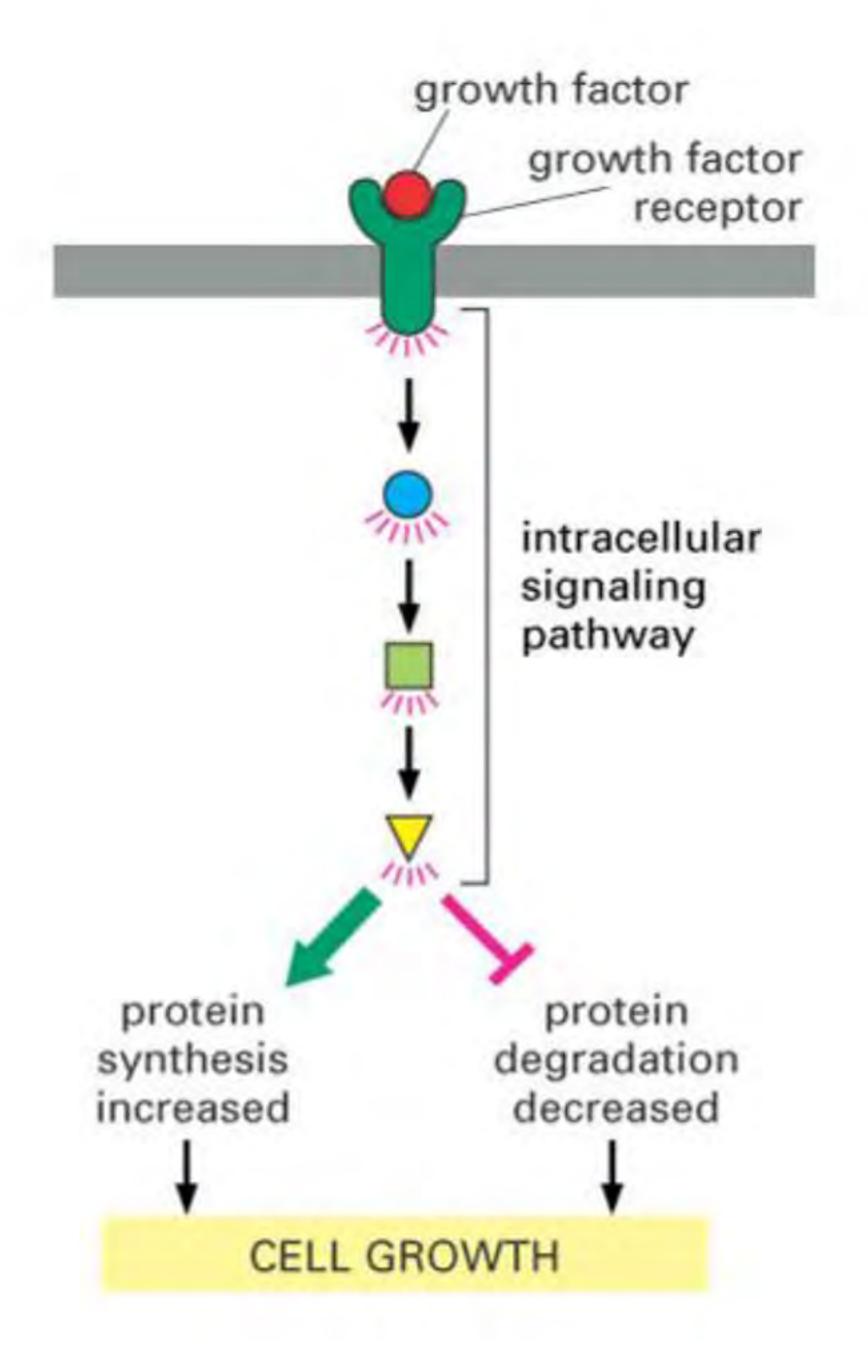
Pathgula	thguide» the pathway resource list			Unelss otherwise noted, the values represent the number or resources that contain/satisfy the specified data/criteria.  The "Contents" section also includes values denoted as "Totals". This is a simple summation and does not account for any overlaps in data between			
lovigation	Complete Listing of All Pathguide Resources	News	General		Organisms		
rotein-Protein nteractions letabolic Pathways	Pathguide contains information about <b>547</b> 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	Major new update		547 277 258	Homo sapiens Mus musculus Saccharomyces cerevisiae Rattus norvegcus	1	
ignaling Pathways athway Diagrams ranscription Factors / ene Regulatory etworks	If you know of a pathway resource that is not listed here, or have other questions or comments, please send us an e-mail.	Visual navigation 2010 Click the 'Databas link on the left mer	Not currently available Free to academic users	Bos taurus 437 Danio rerio 351 Helicobacter pylori 344 Plasmodium falciparum 39 Bacillus subtilis 36 Xenopus laevis 26 Zea mays 8 Human immunodeficience	Arabidopsis thaliana Caenorhabditis elegans Escherichia coli Oryza sativa Schizosaccharomyces pombe		
rotein-Compound nteractions enetic Interaction	<del></del>		Keyword searches Browsing / Canned queries Download in other format Download in BioPAX format Download in PSI format Download in SBML format SQL queries		Danio rerio Helicobacter pylori Plasmodium falciparum Bacillus subtilis Xenopus laevis Zea mays Human immunodeficiency virus		
etworks rotein Sequence ocused other	Protein-Protein Interactions  Database Name (Order: alphabetically   by web popularity )	Full Record					
earch	2P2Idb - The Protein-Protein Interaction Inhibition Database 3D-Interologs - 3D-Interologs	Details Details	Download in CellML format TOOIS  Pathway diagram tools	84	Hepatitis C virus Chlamydomonas reinhardtii Dictyostelium discoideum Mycoplasma pneumoniae		
√ailability	3DID - 3D interacting domains  ADAN - Prediction of protein-protein interaction of modular domains  AHD2.0 - Arabidopsis Hormone Database 2.0	Details Details Details	Contents			- ))	
andards	AllFuse - Functional Associations of Proteins in Complete Genomes  aMAZE - Protein Function and Biochemical Pathways Project	Details Details	Resources containing small Total small molecules	molecule	S	33,680,655	
Reset Search	ANAP - Arabidopsis Network Analysis Pipeline AnimalTFDB - Animal Transcription Factor Database AntiJen - AntiJen a Kinetic, Thermodynamic and Cellular Database	Details Details Details	Resources containing gene Total genes/proteins	s/proteins		277 345,626,442	
atistics atabase Interactions	APID - Agile Protein Interaction DataAnalyzer AS-ALPS - Alternative Splicing - induced ALteration of Protein Structure ASD - Allosteric Database	Details Details Details	Resources containing intera Total interactions/reactions		261 1,311,363,048		
Contact Comments, Questions, Suggestions are Always	ASEdb - Alanine Scanning Energetics Database  ASPD - Artificial Selected Proteins/Peptides Database  ATDB - Animal Toxin Database	Details Details Details	Resources containing pathways Total pathways		2,533,510		
elcome!	AtPID - Arabidopsis thaliana Protein Interactome Database  AtPIN - Arabidopsis thaliana Protein Interactome Network	Details Details	Resources containing exper Total experiments	riments		2,547,553	
	Bacteriome.org - Bacterial Protein Interaction Database for Escherichia Coli	Details			товнекранична	2,01	

STOP 1:
Pathway Profiling /
Compound Mapping

A representation of
a peak (biochemical data)
being associated with
a compound in a database,
then with a compound in
a biological pathway map

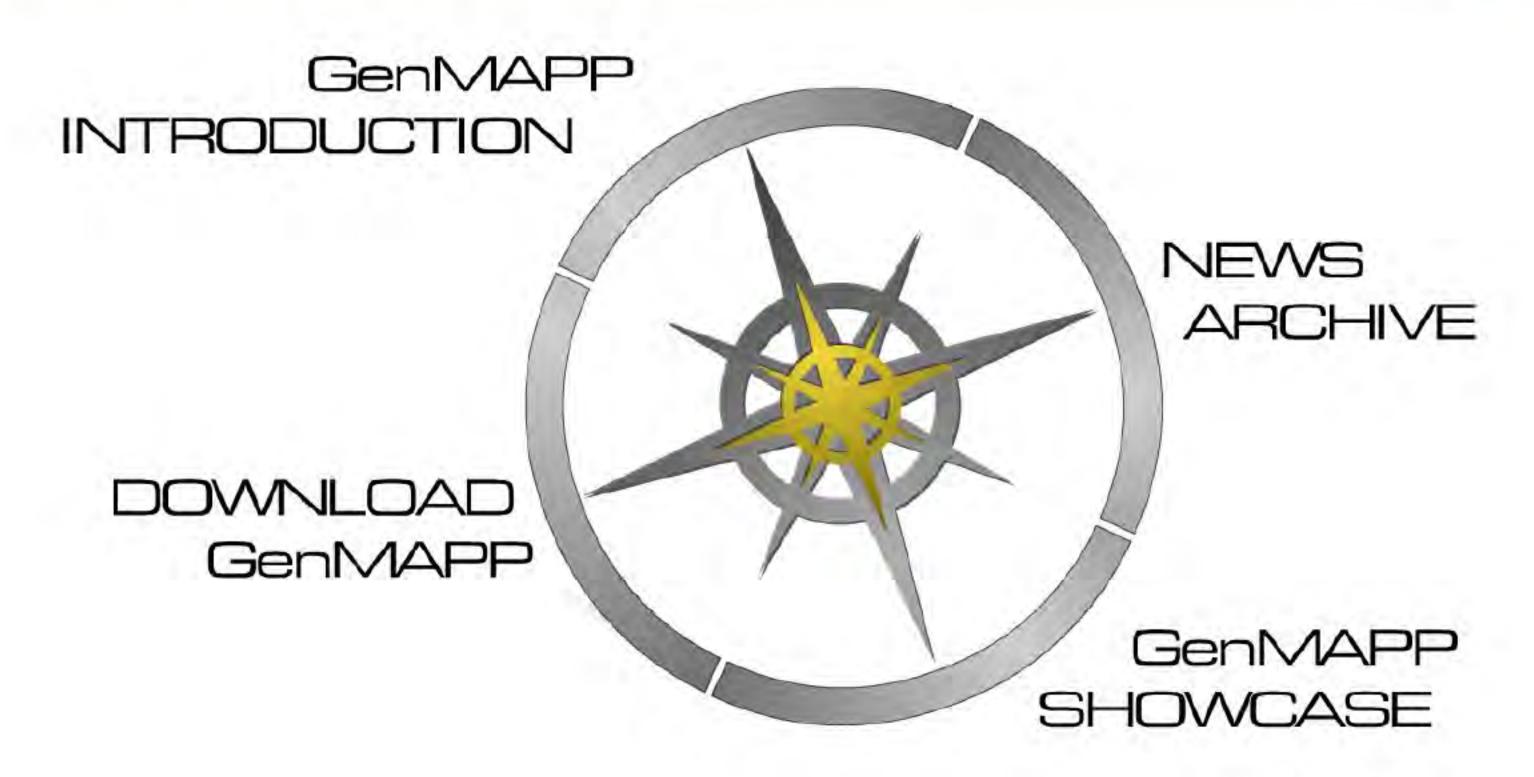


# Biological Pathways/ Signaling Pathways



#### GenMAPP 2.1

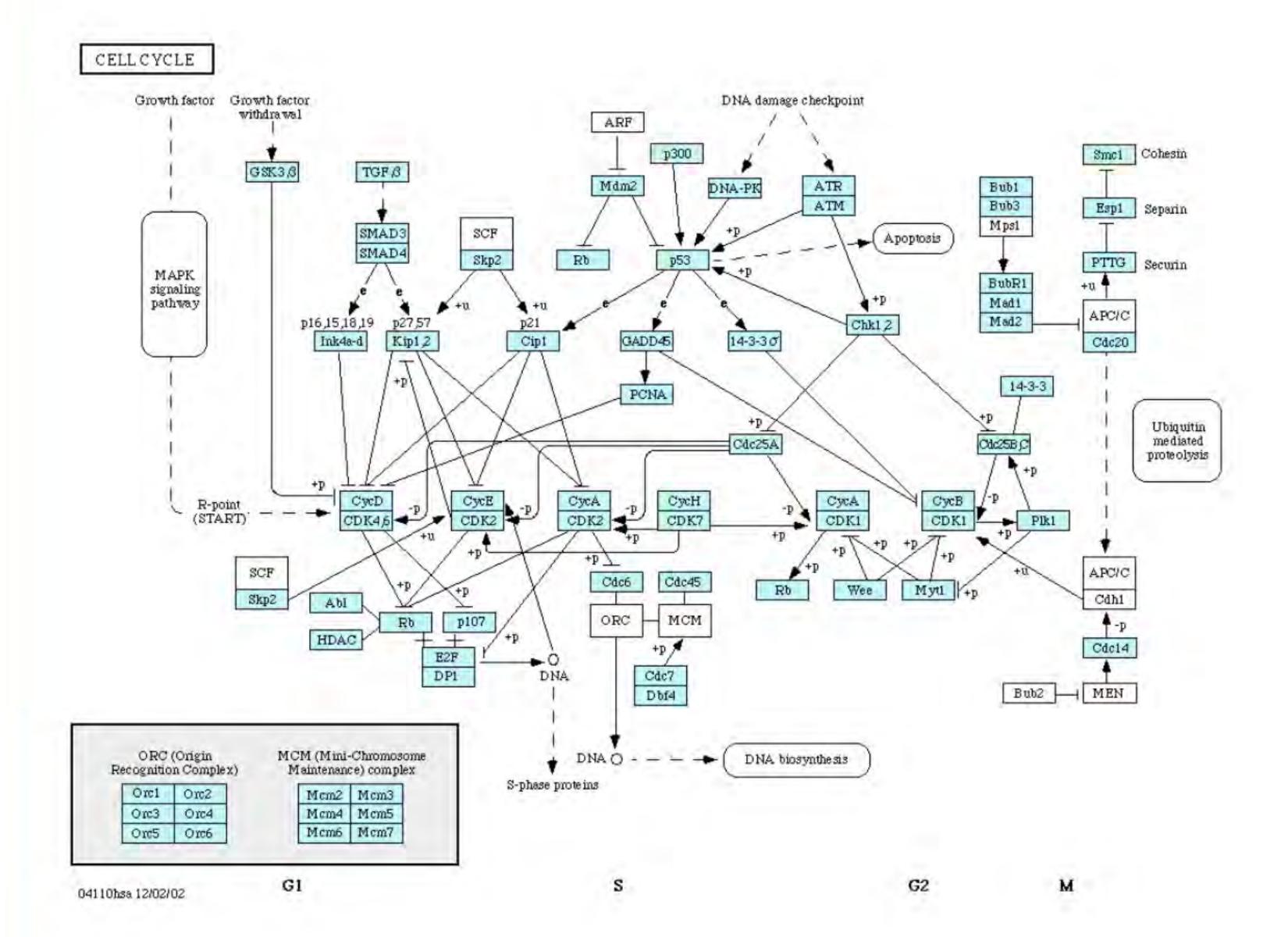




# Pathways for Cellular Processes: Cell Cycle

KEGG Pathways Database on-line:

http://www.genome.ad.jp/kegg



# Signaling Pathways Examples

http://stke.sciencemag.org/content/sigtrans/suppl/2002/10/22/CMP\_7966.DC1/waj ant530.swf

http://stke.sciencemag.org/content/sigtrans/suppl/2002/10/22/CMP\_7966.DC1/waj ant530.swf





#### www.stke.org

Cell Death Mediated by Fas

H. Wajant, Fas Signaling Pathway. (Connections Map, as seen May 2002), http://stke.sciencemag.org/cgi/cm/CMP\_7966

Animation by : Cameron Slayden





www.stke.org

The Jak-STAT Pathway Stimulated by Interferon Y

D. Aaronson, C. M. Horvath, Interferon gamma Pathway. (Connections Maps, as seen May 2002), http://stke.sciencemag.org/cgi/cm/CMP\_9590

Animation by: Carin Cain



KEGG ▼ Search Help

#### **KEGG Home**

Release notes Current statistics Plea from KEGG

#### **KEGG Database**

KEGG overview Searching KEGG KEGG mapping Color codes

#### **KEGG Objects**

Pathway maps Brite hierarchies

#### **KEGG Software**

KegTools KEGG API KGML

#### KEGG FTP

Subscription

GenomeNet

DBGET/LinkDB

Feedback

Kanehisa Labs

#### **Current Statistics**

KEGG Database a Systems informati		
KEGG PATHWAY	Pathway maps, reference (total)	468 (346,709
KEGG BRITE	Functional hierarchies, reference (total)	160 (120,942
KEGG MODULE	KEGG modules, reference (total)	653 (279,875
Genomic informati	on .	
KEGG ORTHOLOGY	KEGG Orthology (KO) groups	18,384
KEGG GENOME	KEGG Organisms	3,565
KEGG GENES	Genes in high-quality genomes (303 eukaryotes, 3057 bacteria, 179 archaea)	15,768,634
KEGG SSDB	Best hit relations within GENES Bi-directional best hit relations within GENES	82,423,307,159 5,010,380,559
KEGG DGENES	Genes in draft genomes (26 eukaryotes)	654,88
KEGG MGENES	Genes in metagenomes (796 samples)	131,900,98
Chemical informat	ion	
KEGG COMPOUND	Metabolites and other small molecules	17,36
KEGG GLYCAN	Glycans	10,98
KEGG REACTION	Biochemical reactions	9,79
KEGG RPAIR	Reactant pair chemical transformations	14,87
KEGG RCLASS	Reaction class	2,94
KEGG ENZYME	Enzyme nomenclature	6,41
Health information	1	
KEGG DISEASE	Human diseases	1,40
KEGG DRUG	Drugs	10,14
KEGG DGROUP	Drug groups	1,74
KEGG ENVIRON	Crude drugs and health-related substances	85

# KEGG Atlas: Cancer Pathways



Example: hsa05200 Pathways in cancer

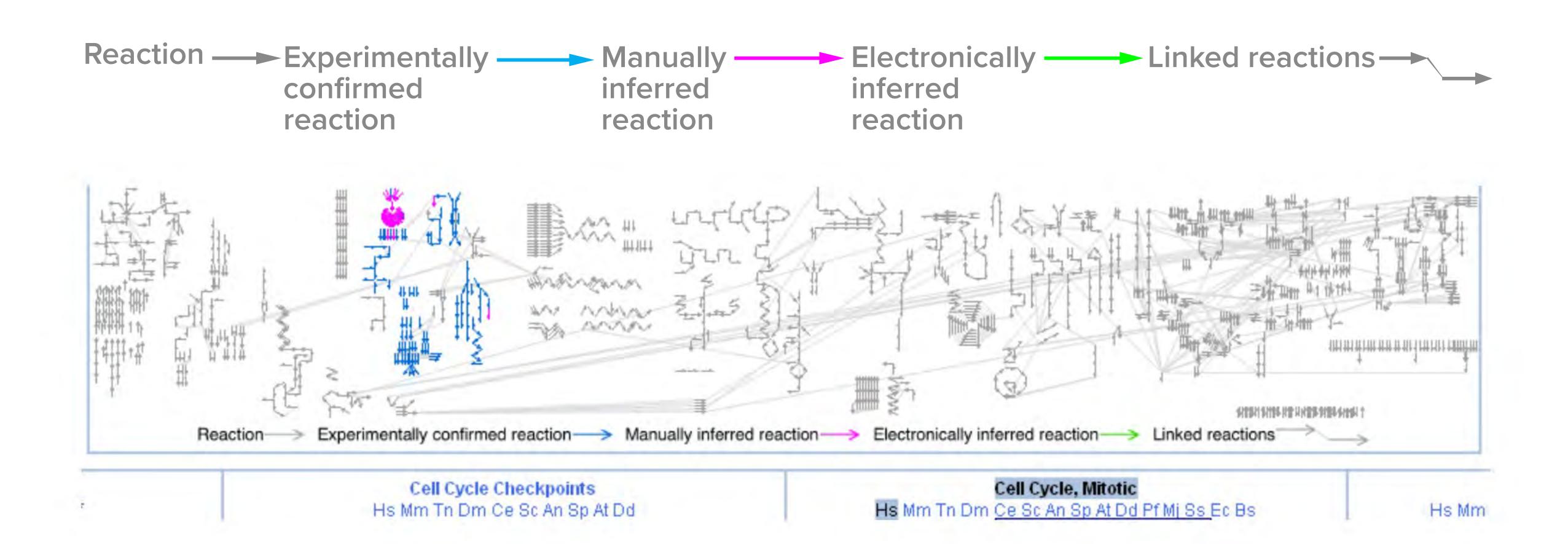
# Reactome - a knowledgebase of biological processes

The Reactome project is a collaboration among Cold Spring Harbor Laboratory, The European Bioinformatics Institute, and The Gene Ontology Consortium to develop a curated resource of core pathways and reactions in human biology.

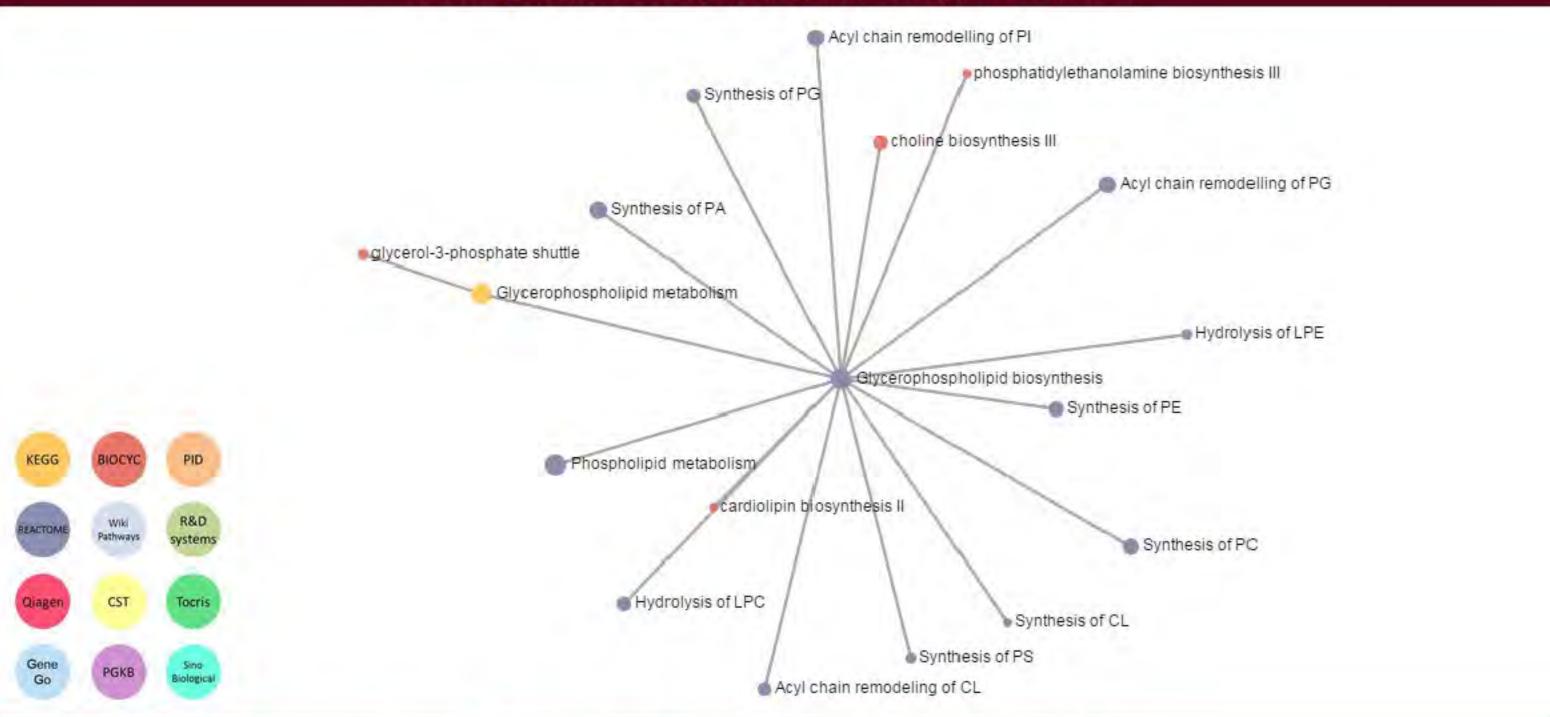
The information in this database is authored by biological researchers with expertise in their field, maintained by the Reactome editorial staff, and cross-referenced with PubMed, GO, and the sequence databases at NCBI, Ensembl and UniProt.



## Reactome - a knowledgebase of biological processes



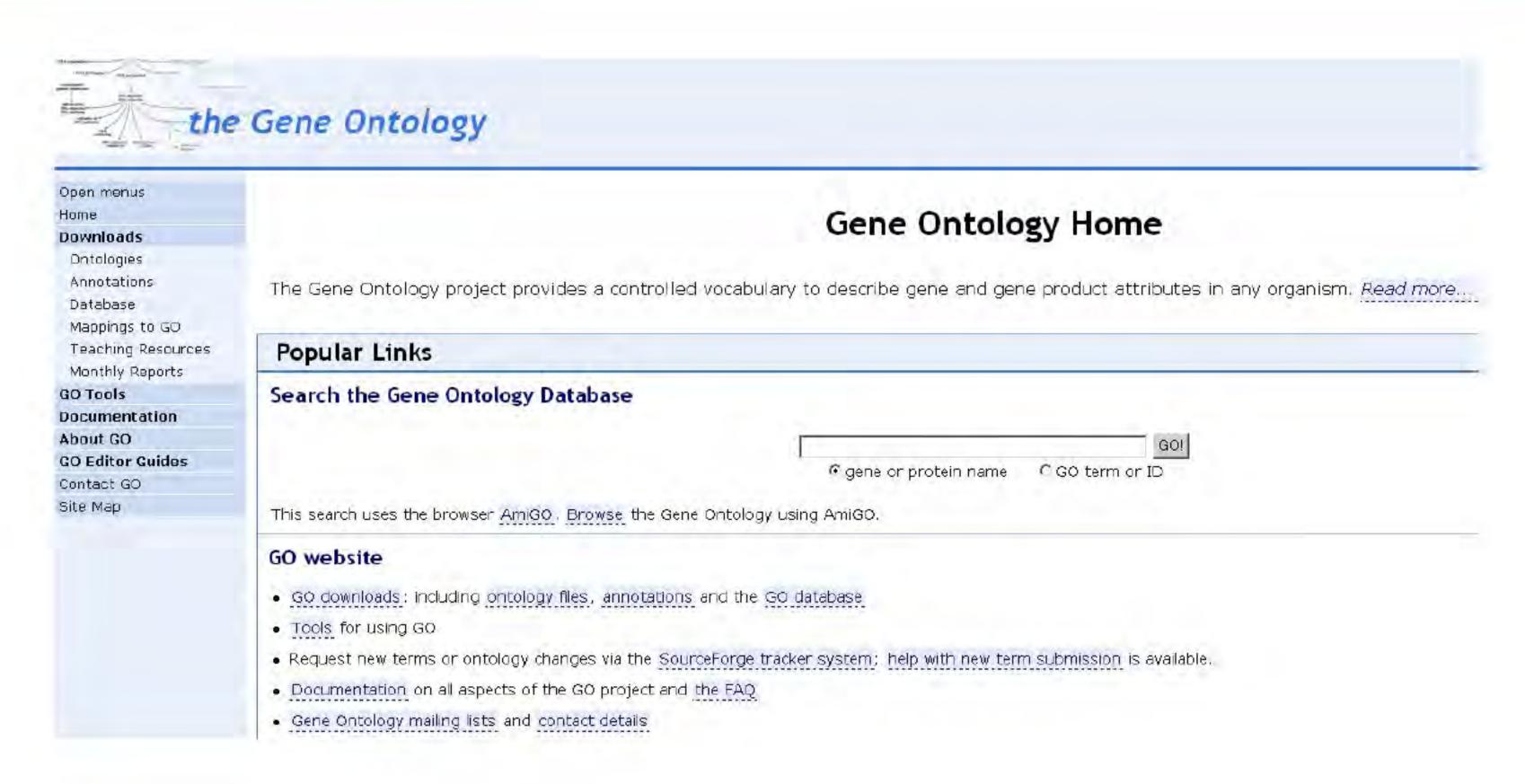




18 Pathways in the Glycerophospholipid biosynthesis SuperPath

http://pathcards.genecards.org/card/glycerophospholipid\_biosynthesis

## STOP 2: Beyond Pathways: FUNCTIONAL CLASSIFICATION using Gene Ontologies (GO)



http://www.geneontology.org/

# The Gene Ontologies

A Common Language for Annotation of Genes from Yeast, Flies and Mice

...and Plants and Worms

...and Humans

...and anything else!

### Gene Ontology Objectives

# GO represents concepts used to classify specific parts of our biological knowledge

Biological Process

Molecular Function

Cellular Component

GO develops a common language applicable to any organism

GO terms can be used to annotate gene products from any species, allowing comparison of information across species

## GO Three (Orthogonal) Ontologies

#### Molecular Function

elemental activity or task e.g. DNA binding, catalysis of a reaction

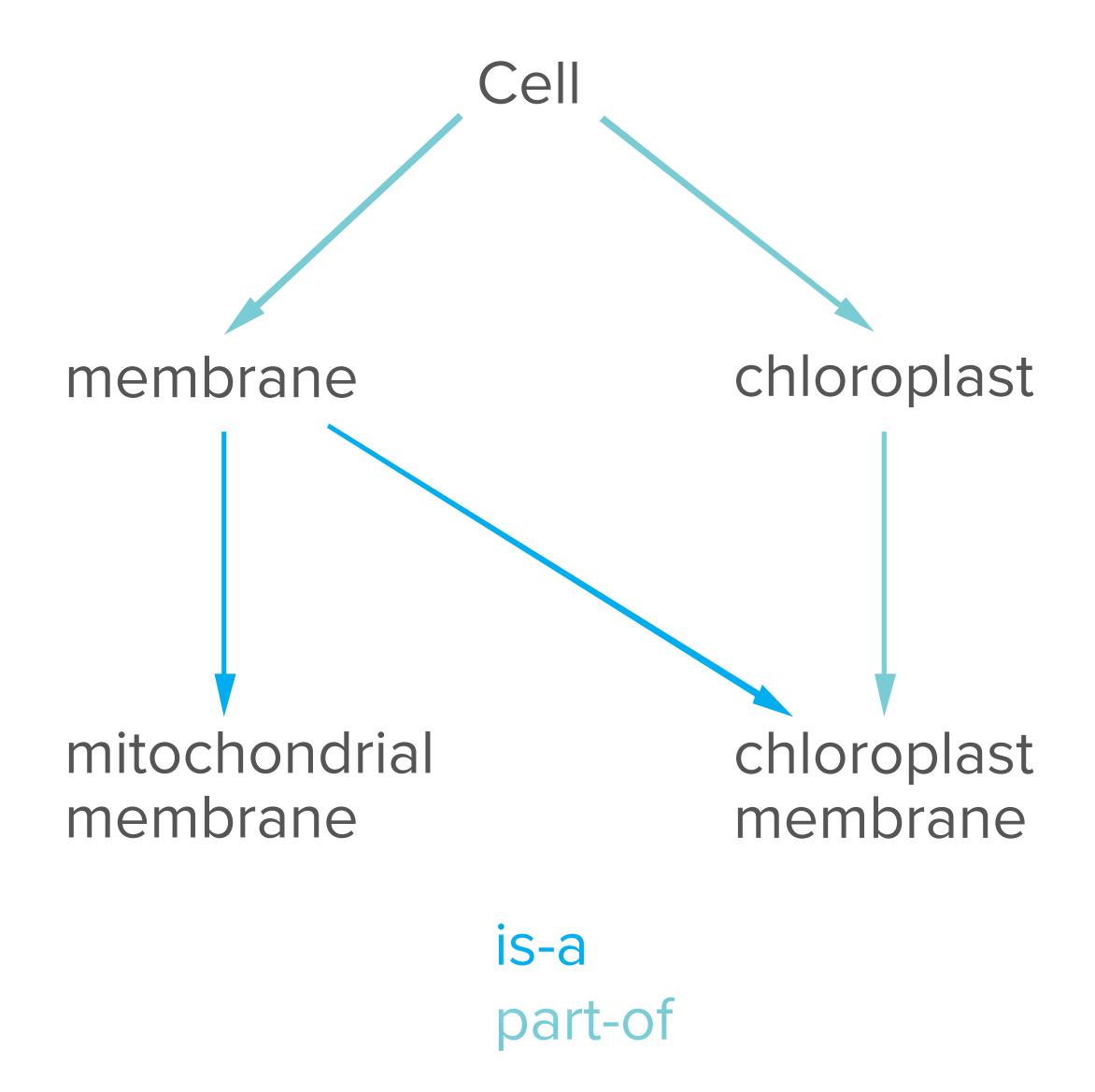
#### **Biological Process**

broad objective or goal e.g. mitosis, signal transduction, metabolism

#### Cellular Component

location or complex e.g. nucleus, ribosome

#### Directed Acyclic Graph



#### Example: Gene Product = Hammer

Function (what)

Drive nail (into wood)

Drive stake (into soil)

Smash roach

Clown'a juggling object



Process (why)

Carpentry

Gardening

Pest Control

Entertainment

#### What can scientists do with GO?

#### Access gene product functional information

Provide a link between biological knowledge, gene expression profiles and proteomics data

Find how much of a proteome is involved in a process/ function/component in the cell

using a GO-Slim

(a slimmed down version of GO to summarize biological attributes of a proteome)

# Map GO terms and incorporate manual GOA annotation into own databases

to enhance your dataset

or to validate automated ways of deriving information about gene function (text-mining)

### Access to the Gene Ontology

#### Downloads

```
formats available:
```

OBO GO

XML OWL

MySQL

(http://www.geneontology.org/GO.downloads)

#### Web-based tools

AmiGO

(http://www.godatabase.org)

QuickGO

(http://www.ebi.ac.uk/ego)

#### GO-based Tools

Functional profiling of microarray and proteomics data

#### Tools for gene expression/microarray analysis

BiNGO GeneMerge

CLENCH GFINDer: Genome Function

DAVID GOArray

EASE GOdist

eGOn v2.0 GOODIES

ermineJ GoMiner and MatchMiner

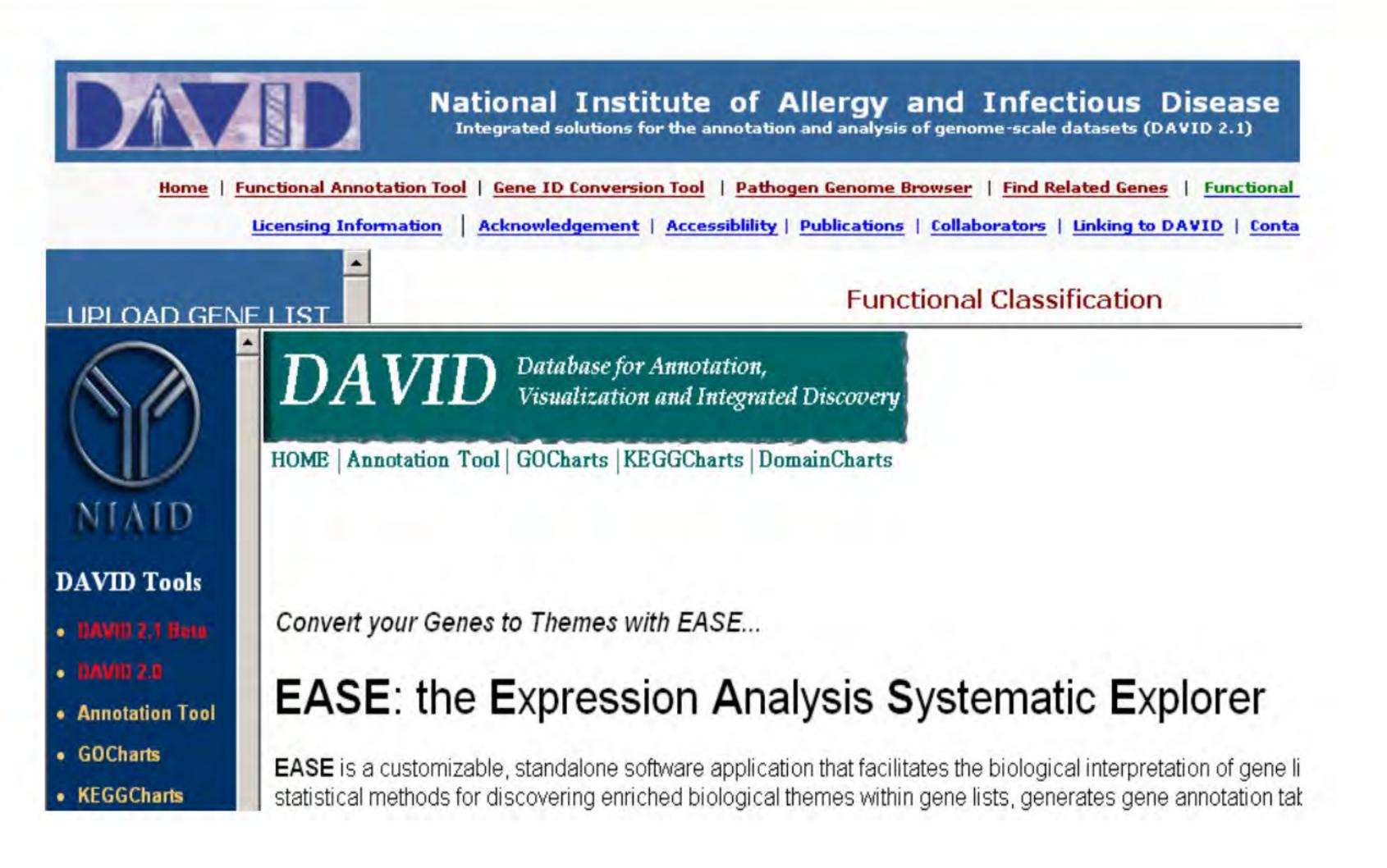
FatiGO GOstat

FuncAssociate GoSurfer

FuncExpression GO Term Finder

GARBAN GOTM (Gene Ontology Tree Machine)

# Combining Pathways and GO ontology enrichment analysis



#### **DAVID**

Database for Annotation, Visualisation and Integrated Discovery

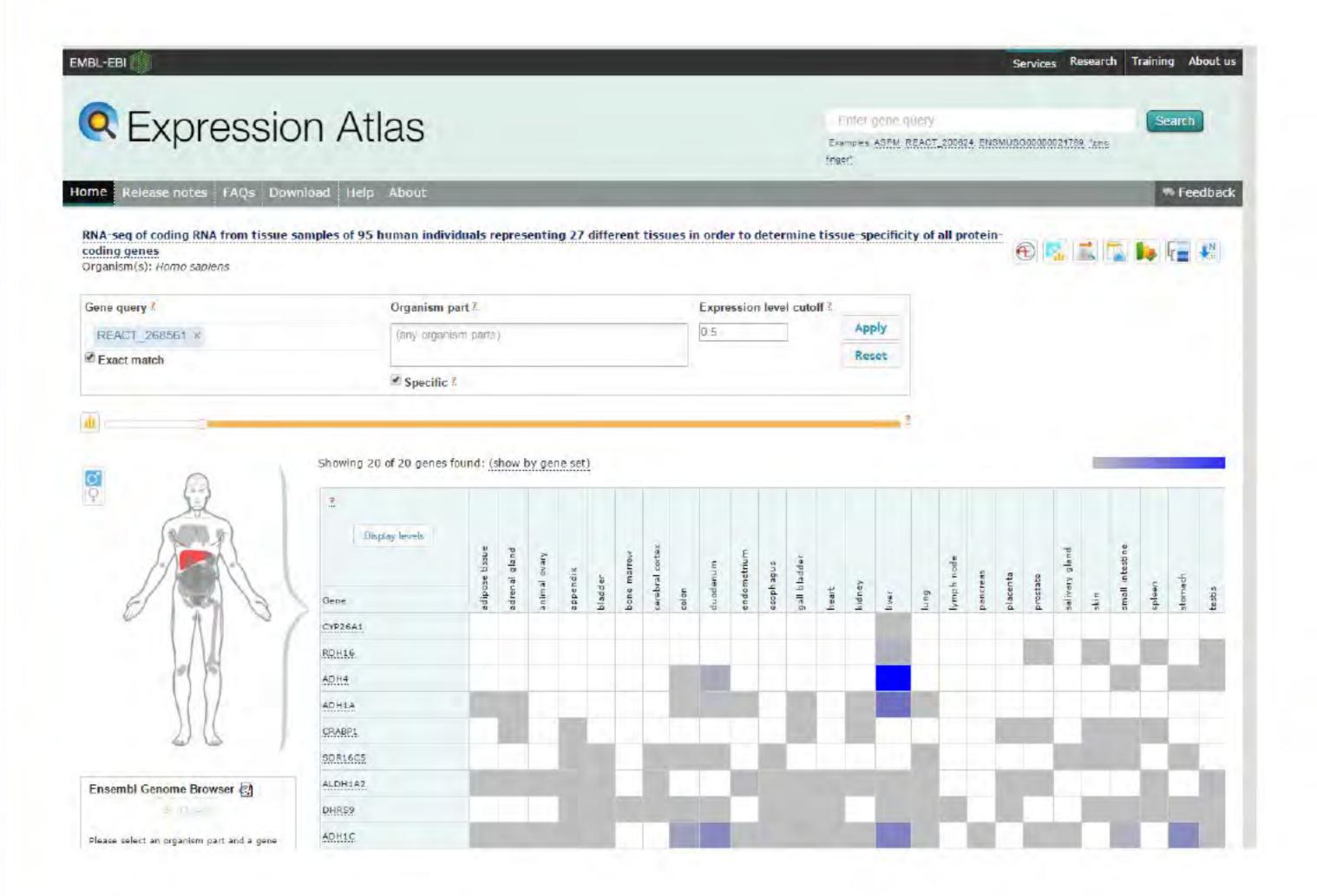
### NEXT STOP: Next Generation of Systems Biology Tools

Bringing Tools and Data together

Open Access Web-based applications

### Data Resources on-line: Expression Atlas (EBI)

http://www.ebi.ac.uk/gxa//experiments/E-MTAB-1733?gene-Query=REACT\_268561&serialized FilterFactors



# Pathway and Network analysis on-line: Public Domain

http://www.pathwaycommons.org/about/
http://www.ndexbio.org/
http://amp.pharm.mssm.edu/Enrichr/
CellMiner: http://discover.nci.nih.gov/cellminer/



#### For biologists

Search, visualize and download Pathway Commons pathways as part of an integrated network analysis (more)











# Systems Biology and Big Data: connecting tools to data (hands-on)

#### GEO2Enrichr

Google Chrome Extension to extract gene sets from GEO and analyze using Systems

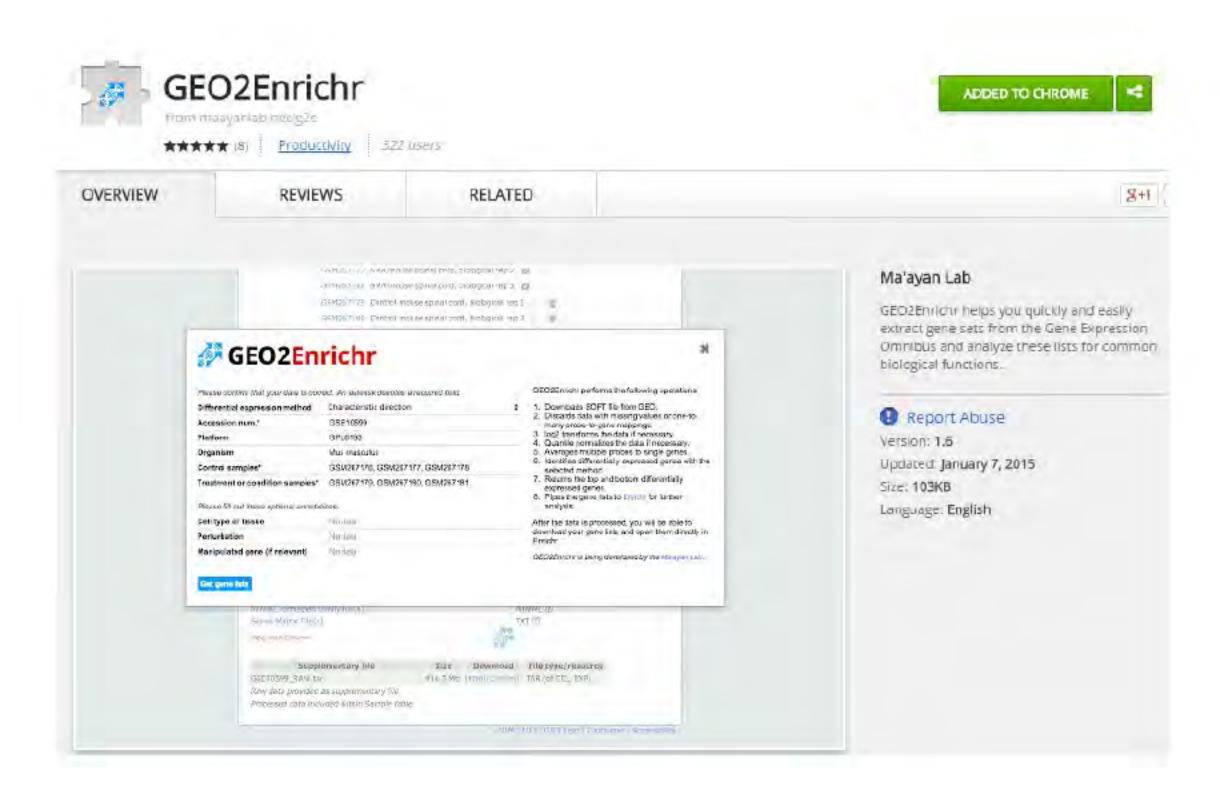
Biology Tools on-line

#### Data Public Repositories

Gene Omnibus - GEO Express Array Short Reads Archive - SRA

#### Systems Biology Tools

Pathway Analysis Gene Ontology Network Analysis



#### Enrichr



Login | Register

102799 lists analyzed!

What's New? Dataset Statistics Find A Gene About Help

Input data

Choose an input file to upload. Either in BED format or a list of genes. For a list of genes: separate each gene symbol with a new line. For a quantitative set, add a comma and the level of membership of that gene between 0 and 1 after each gene symbol. Try an example BED file.

Choose File No file chosen

Or paste in a list of gene symbols optionally followed by a comma and levels of membership between 0 and 1 with each gene separated by a new line. Try a regular example or an example of a quantitative set.

Mrpl9 LOC100046168 Zfp11 Asf1a Plscr2 Mettl7a Rab1 Lrrc19 LOC100047214 Phf7

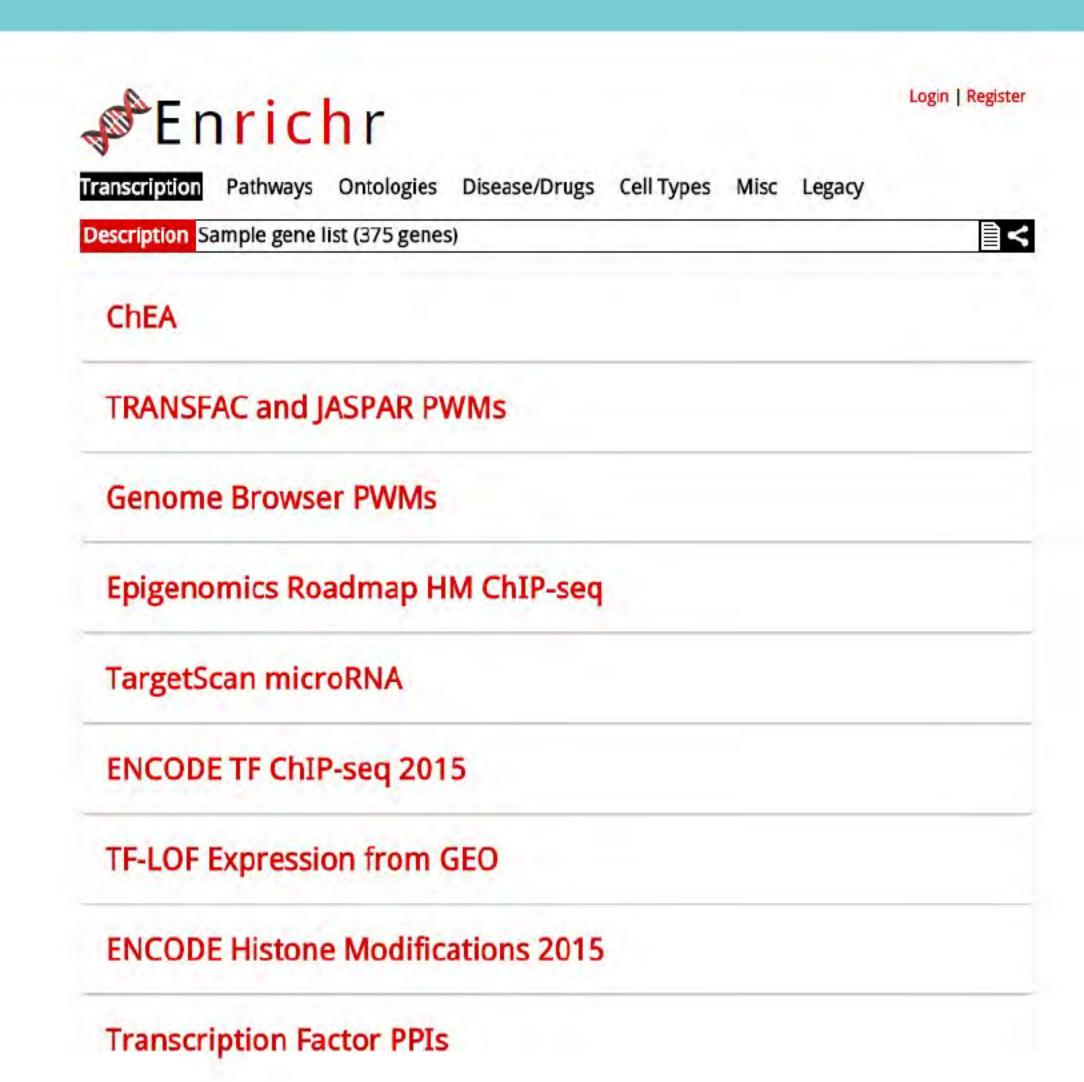
0 gene(s) entered

Enter a brief description for the list in case you want to share it. (Optional)

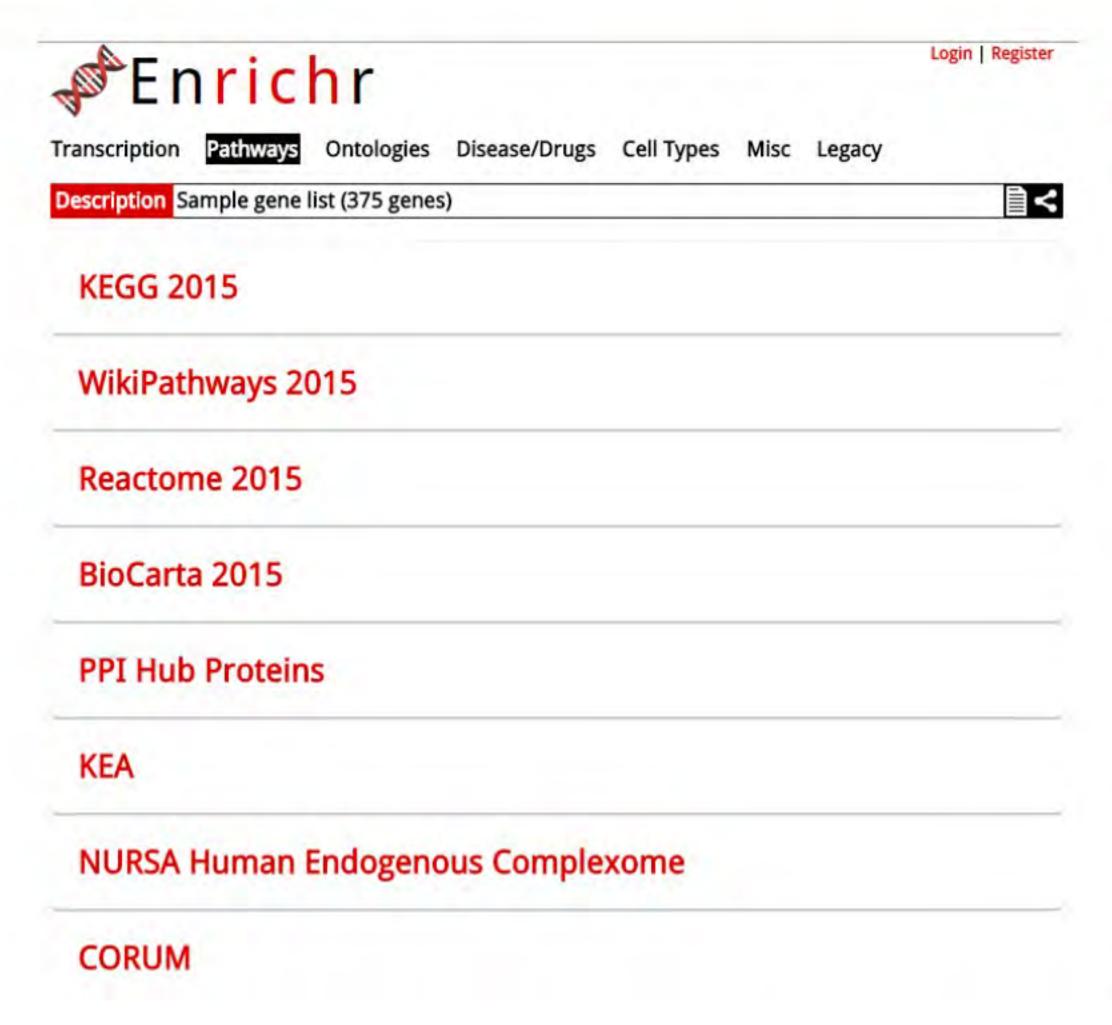
Sample gene list

Please acknowledge Enrichr in your publications by citing the following reference: Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics, 2013;128(14).





#### Enrichr



Panther

10

Bar Graph

Search:





Hover each row to see the overlapping genes.

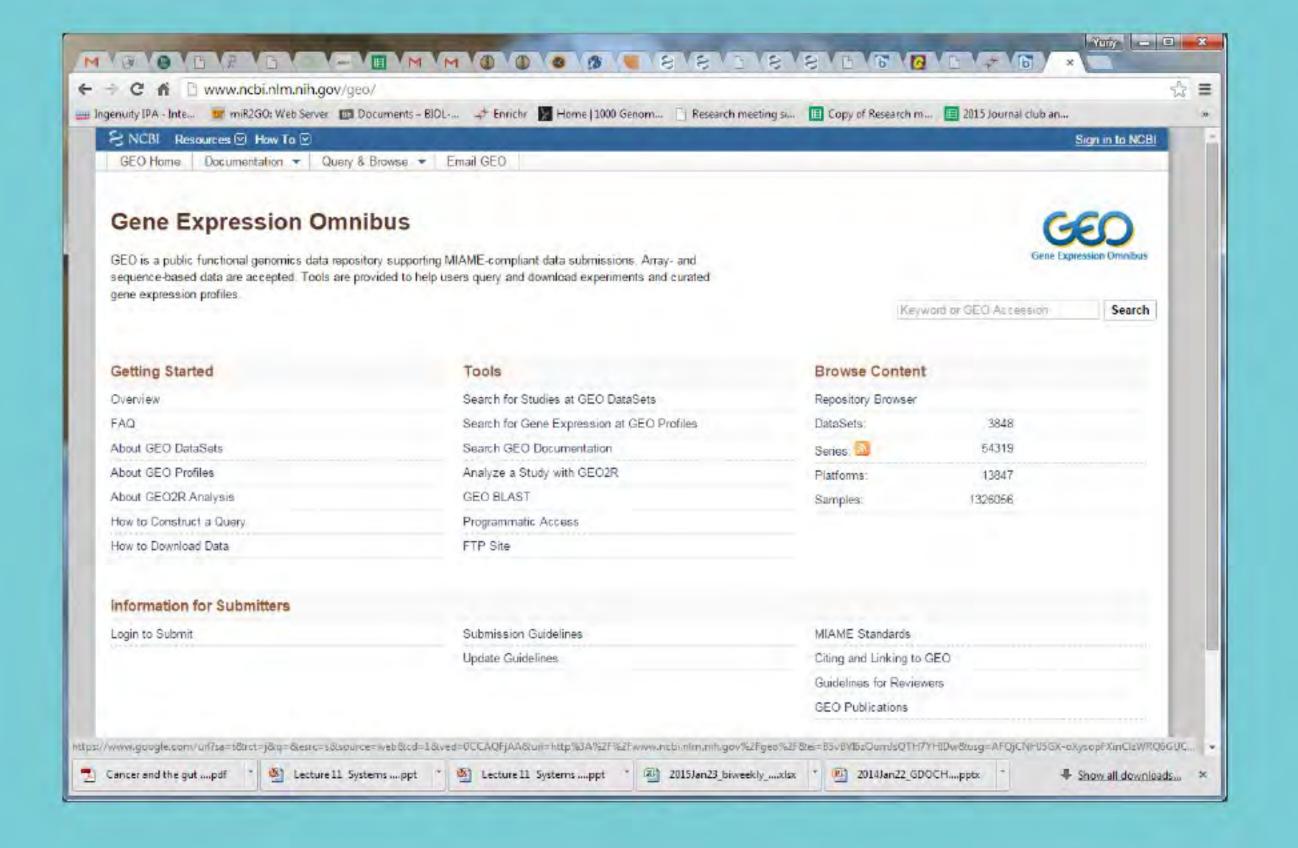
entries per page

Combined Z-P-Index Name value Score \* score Beta1 adrenergic receptor signaling pathway\* 0.02936 1.29 -1.80Purine metabolism\* 0.01387 -1.501.07 Metabotropic glutamate receptor group II 0.07184 -1.370.88 pathway\* Thyrotropin-releasing hormone receptor signaling 4 0.72 0.1257 -1.12pathway\* Muscarinic acetylcholine receptor 2 and 4 signaling 5 0.1114 -1.090.70 pathway\* Heme biosynthesis\* 0.02116 -0.890.64 Insulin/IGF pathway-mitogen activated protein 0.07184 -0.910.58 kinase kinase/MAP kinase cascade\* Beta2 adrenergic receptor signaling pathway\* 0.1352 0.00 1.61 Coenzyme A biosynthesis\* 0.1352 3.69 0.00 Beta3 adrenergic receptor signaling pathway\* 0.1601 0.00 2.51

Showing 1 to 10 of 50 entries | Export entries to table Terms marked with an \* have an overlap of less than 5 Previous Next

#### http://www.ncbi.nlm.nih.gov/geo/

http://www.ncbi.nlm.nih.gov/geo/que-ry/acc.cgi?acc=GSE46141







ME | SEARCH | SITE MAR GEO Publications FAQ MIAME Email GEO NCBI > GEO > Accession Display 2 Not logged in | Login 2 Scope: Self ▼ Format: HTML ▼ Amount: Quick ▼ GEO accession: GSE46141 Series GSE46141 Query DataSets for GSE46141 Status Public on Dec 02, 2013 Title Expression data from fine-needle aspiration biopsies of breast cancer metastases from different anatomical sites Organism Homo sapiens Expression profiling by array Experiment type Breast cancer molecular subtypes preferentially metastasize to specific organs Summary and the anatomical location of the metastasis is associated with the length of survival post-recurrence. We used microarrays to provide a detailed characterization of breast cancer sitespecific metastases with particular focus on identifying genes predictive of breast cancer liver metastatic proprnsity Overall design We performed global gene expression profiling on fine-needle aspirates of metastatic lesions from different anatomical sites obtained from breast cancer patients treated within the Swedish randomized trial (TEX) of first-line chemotherapy for locally advanced or metastatic breast cancer. Samples were collected before commencement of treatment. Contributor(s) Hedenfalk I, Hatschek T Kimbung S, Kovács A, Bendahl PO, Malmström P et al. Claudin-2 is an Citation(s) independent negative prognostic factor in breast cancer and specifically predicts early liver recurrences. Mol Oncol 2014 Feb;8(1):119-28. PMID: 24287398 Apr 17, 2013 Submission date Last update date Jul 09, 2014 Contact name Siker Kimbung Organization name Lund University Department of Oncology and Pathology, Clinical Sciences, Lund Department Lab Lund University Cancer Center/Medicon Village/ Building 404:C2 Street address Scheelevägen 2 Lund ZIP/Postal code SE-22381 Country Sweden GPL10379 Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray Platforms (1) [HuRSTA-2a520709] Samples (91) # More... GSM1124867 KI\_TEX\_2791\_112 liver GSM1124868 KI\_TEX\_2987\_146 liver

GSM1124869 KI\_TEX\_1594\_94 breast

#### http://www.ncbi.nlm.nih.gov/geo/

# http://www.ncbi.nlm.nih.gov/geo/que-ry/acc.cgi?acc=GSE46141

Platforms (1)	GPL10379 Rosetta/Merck Huma [HuRSTA-2a520709]		Custom Affymetri	x 2.0 microarray		
Samples (91)			Ctrl Expmt			
More	GSM1124867 KI_TEX_2791_1:	12 liver	0 0			
	GSM1124868 KI_TEX_2987_14	46 liver	0 0			
	GSM1124869 KI_TEX_1594_94	4 breast	0 0			
Relations						
BioProject	PRJNA197346					
Analyze with G	EO2R					
Download fam	ily		For	mat		
SOFT formatted family file(s)			SOFT 2			
MINIML formatte	ed family file(s)		MIN	IML 2		
Series Matrix File	e(s)		TXT	?		
Pipe into Enrichi						
S	Supplementary file	Size	Download	File type/resource		
GSE46141_Prob	eset_Annotation_20100204.txt.gz	9.3 Mb	(ftp)(http)	TXT		
GSE46141_RAV	/.tar	469.9 Mb	(http)(custom	TAR (of CEL)		
Raw data provid	ed as supplementary file					
Processed data in						

Contributor(s)	Hedenfalk I, Hatschek T						
Citation(s)	Kimbung S, Kovács A, Bendahl PO, Malmström P et al. Claudin-2 is an independent negative prognostic factor in breast cancer and specifically predicts early liver recurrences. <i>Mol Oncol</i> 2014 Feb;8(1):119-28. PMID: 24287398						
Submission date	Apr 17, 2013						
Last update date							
Contact name	Siker Kimbung						
Organization name	Lund University						
Department	Department of Oncology and Pathology, Clinical Science	es, L	und				
Lab	Lund University Cancer Center/Medicon Village/ Buildin	ig 40	04:C2				
Street address	Scheelevägen 2						
City	Lund						
ZIP/Postal code	SE-22381						
Country	Sweden						
Platforms (1)	GPL10379 Rosetta/Merck Human RSTA Custom Affym [HuRSTA-2a520709]	etri	x 2.0 microarray				
Samples (91)	Ctrl Expmt						
∃ Less	GSM1124867 KI_TEX_2791_112 liver						
	GSM1124868 KI_TEX_2987_146 liver 🔲 😿						
	GSM1124869 KI_TEX_1594_94 breast 🕝 📋						
	GSM1124870 KI_TEX_1583_8 breast	1					
	GSM1124871 KI_TEX_1524_35 lymph node						
	GSM1124872 KI_TEX_1467_82 skin local-regional		8				
	GSM1124873 KI_TEX_2904_71 skin local-regional						
	GSM1124874 KI_TEX_2850_62 lymph node						
	GSM1124875 KI_TEX_2917_74 lymph node						
	GSM1124876 KI_TEX_2777_105 skin local-regional						
	GSM1124877 KI_TEX_1451_81 skin local-regional						
	GSM1124878 KI_TEX_1560_10 liver		€				
	GSM1124879 KI TEX 2782 60 skin		EI .				

Questions?

