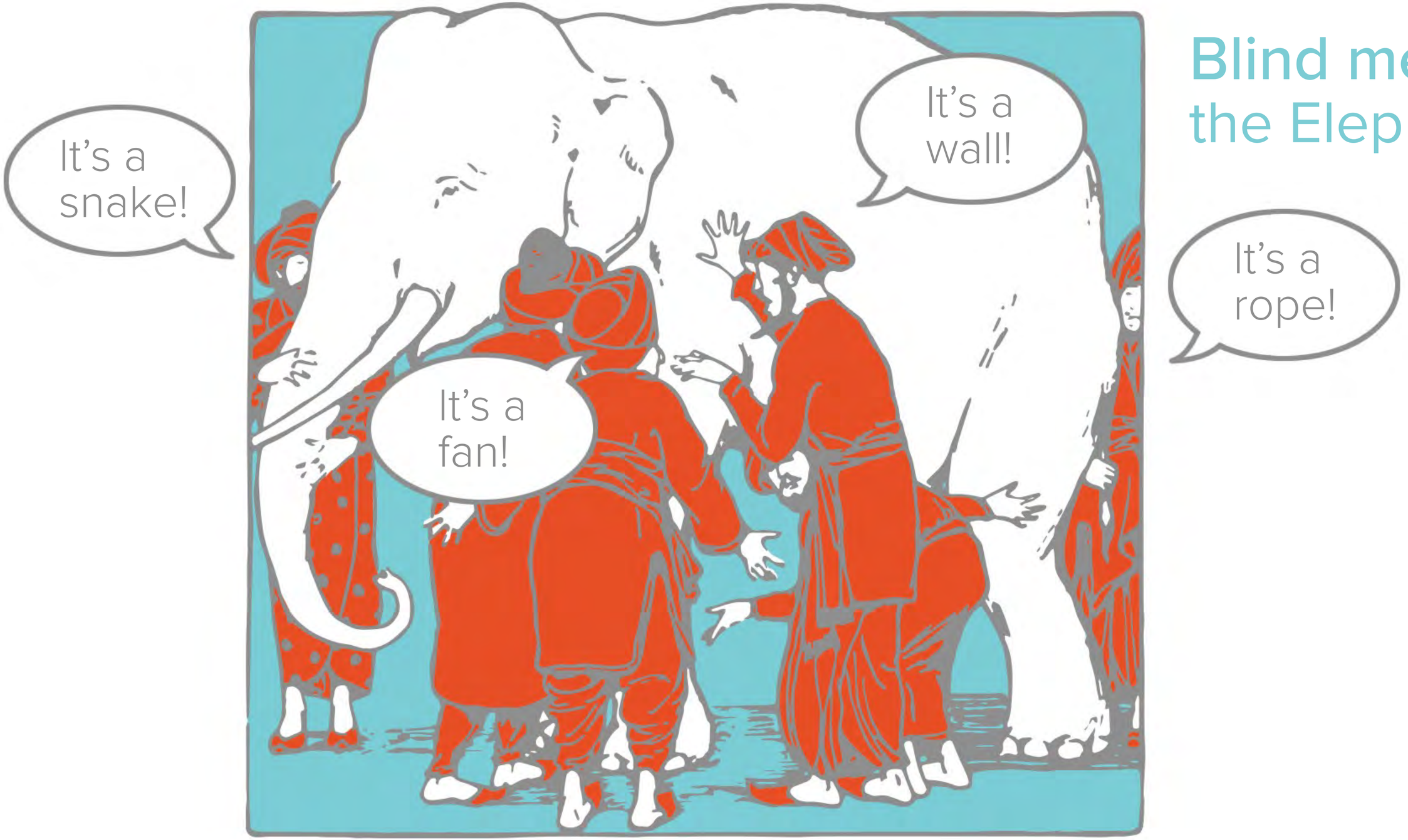
Introduction to Systems Biology

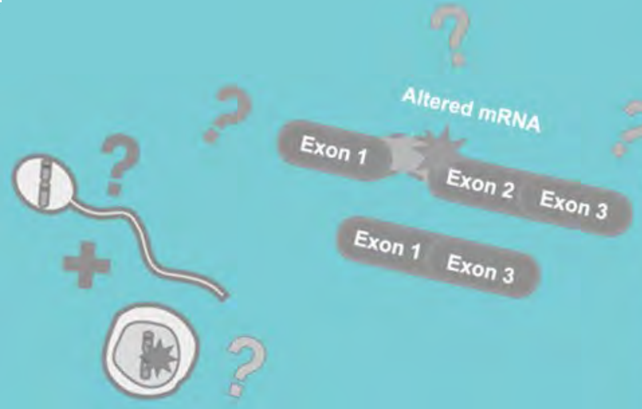
- Definitions and fundamental concepts
- Tools for Systems Biology analysis – overview
- Pathway Analysis: overview
- Gene Ontology Analysis: overview
- Public on-line resources – hands-on

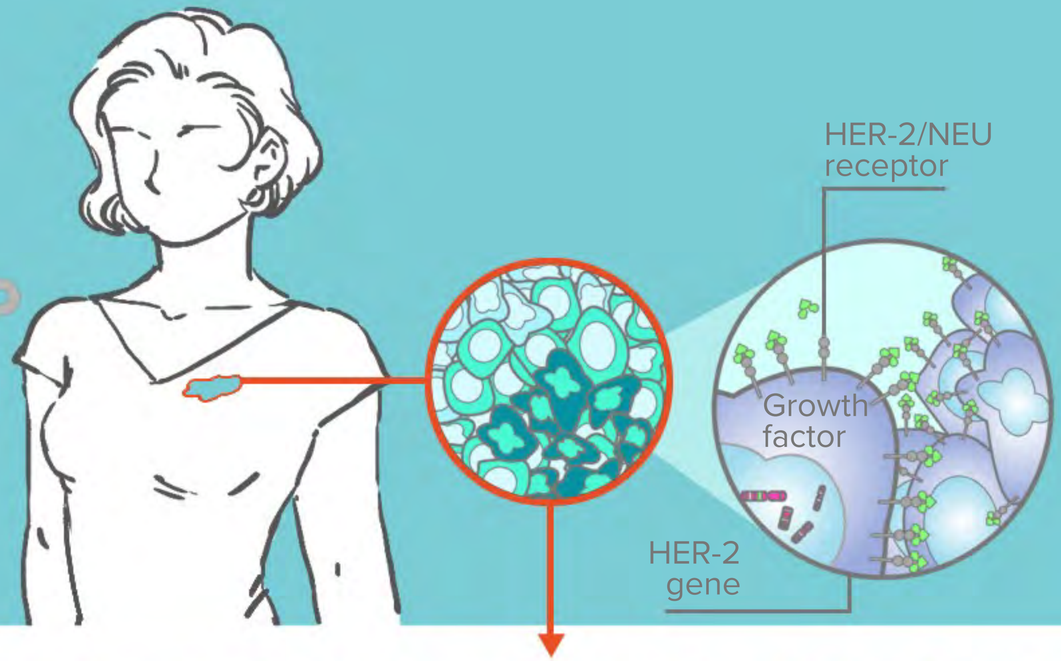
Blind men & the Elephant



Personalized medicine:

A daunting challenge





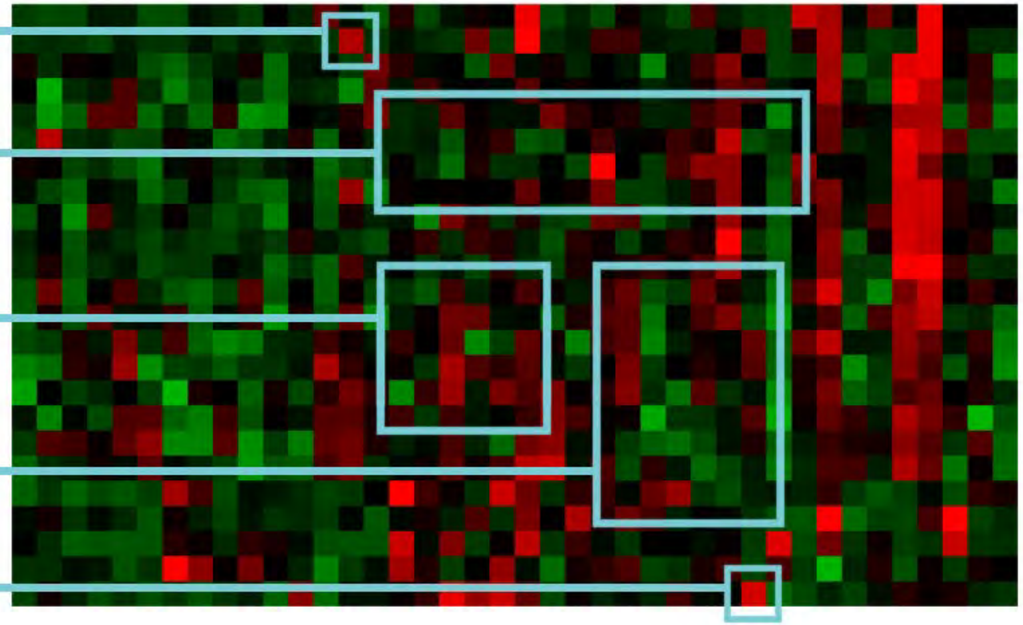
Growth receptor present

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.

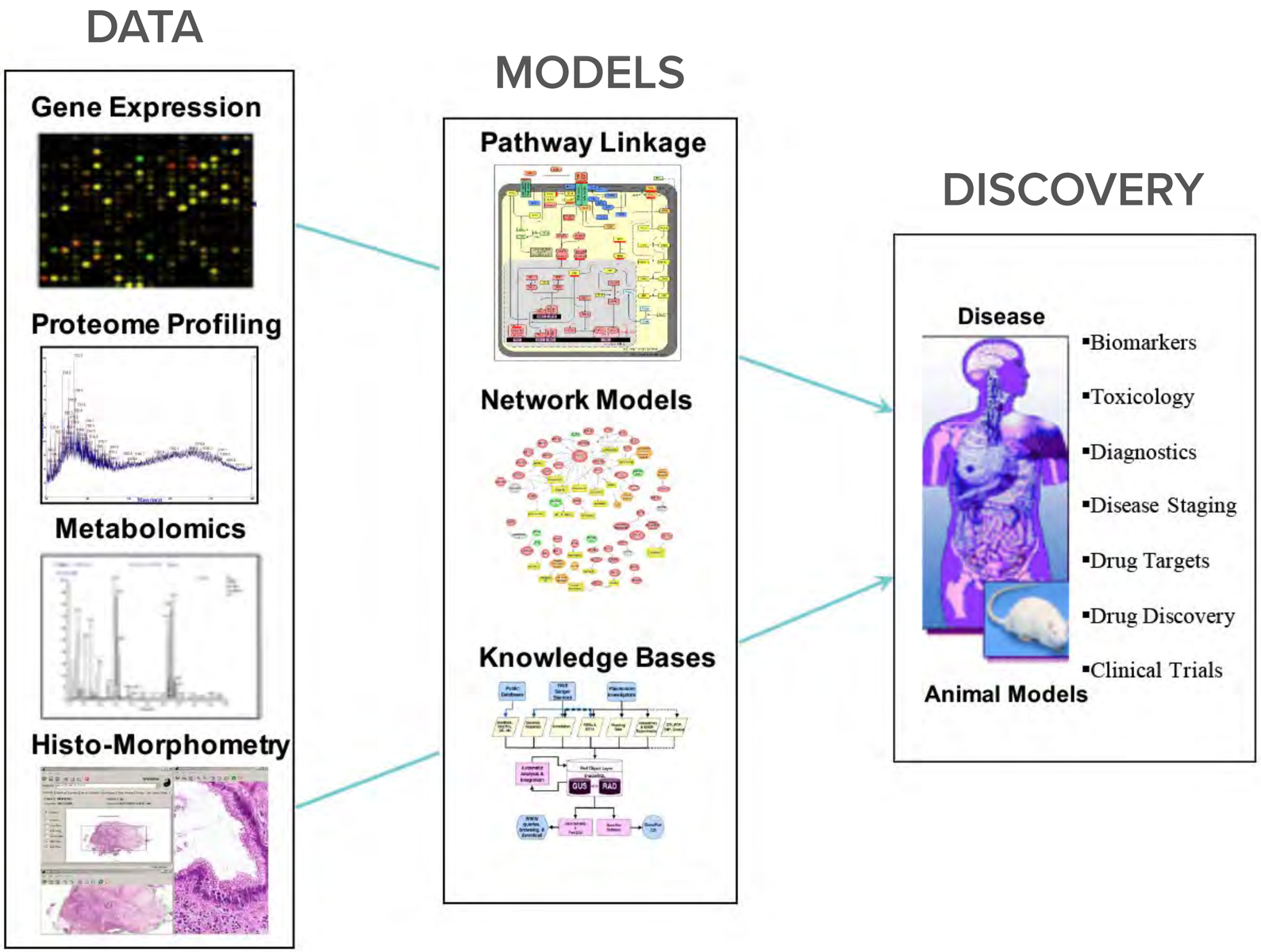
Living systems are dynamic and complex, and their behavior may be hard to predict from the properties of individual parts.

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.

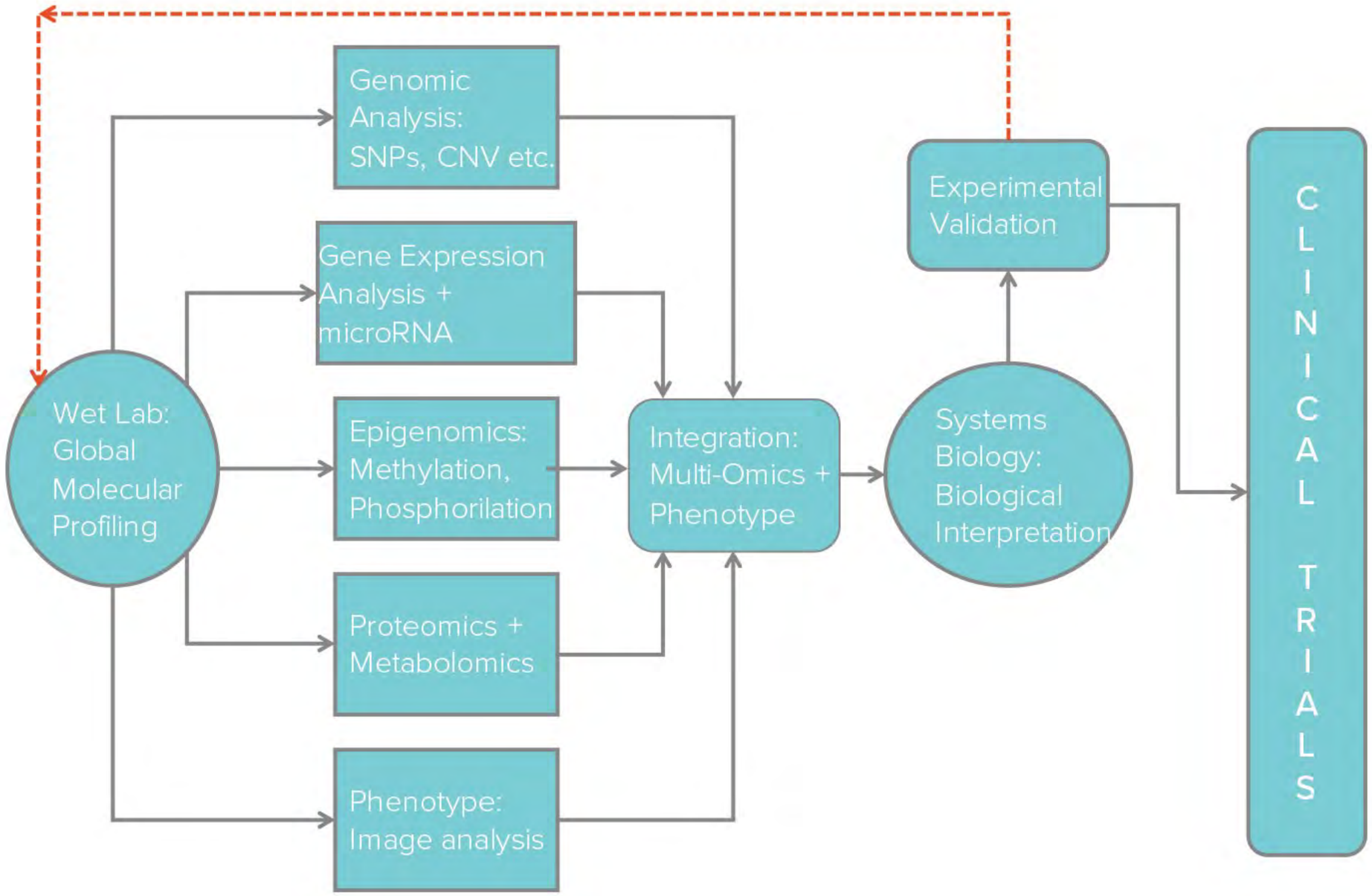
Systems problems are emerging as central to all areas of biology and medicine.

- Dep. Systems Biology, Harvard University

Systems Biology /
Integrative Bioinformatics



How does it all fit together?



Systems Biology Methods/ Analysis tools

Pathway Analysis

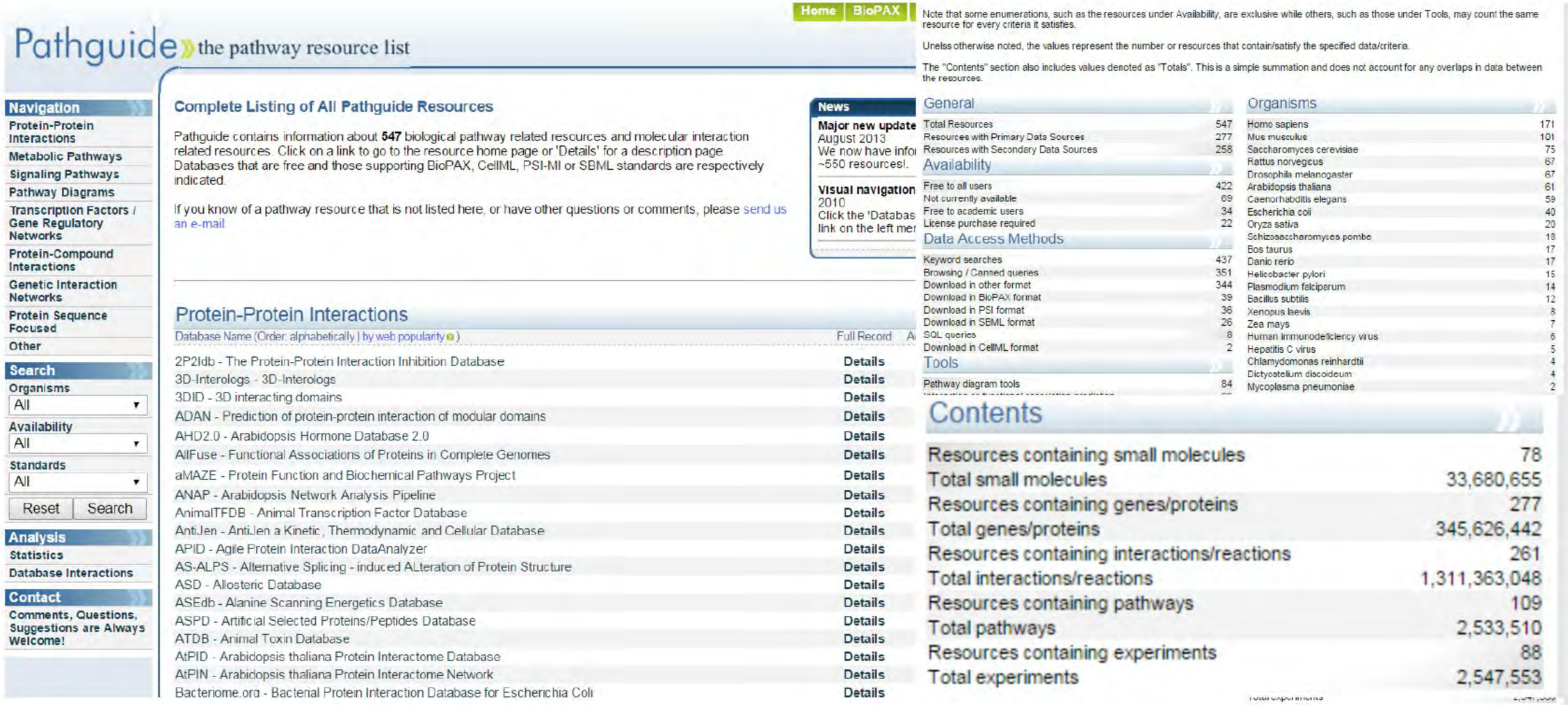
Biological Pathways/ Signaling Pathways - overview
Bioinformatics Tools and Resources for Pathway Analysis
Pathway Studio (Ariadne/Elsevier)
GeneGo (Thomson Scientific/Reuters)
Ingenuity Pathway Analysis IPA (Ingenuity Systems)
Applications: micro-array and other omics- data analysis and text mining

Gene Ontology – GO Enrichment Analysis

Biological Interactions Networks

From Pathways and Networks to Biological Interpretation

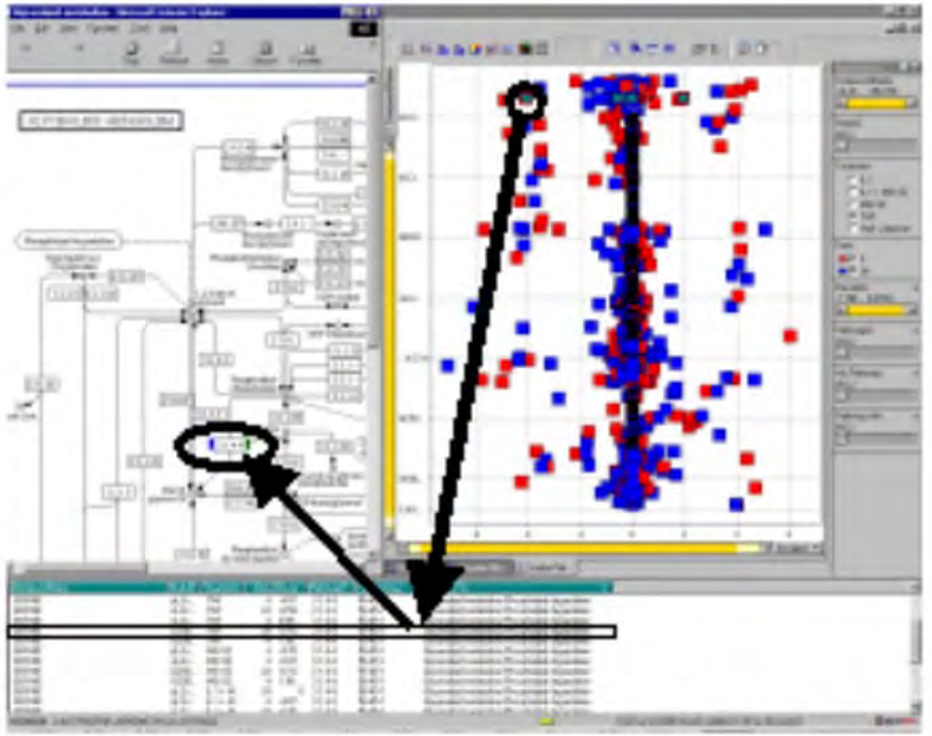
Pathway Resources Central



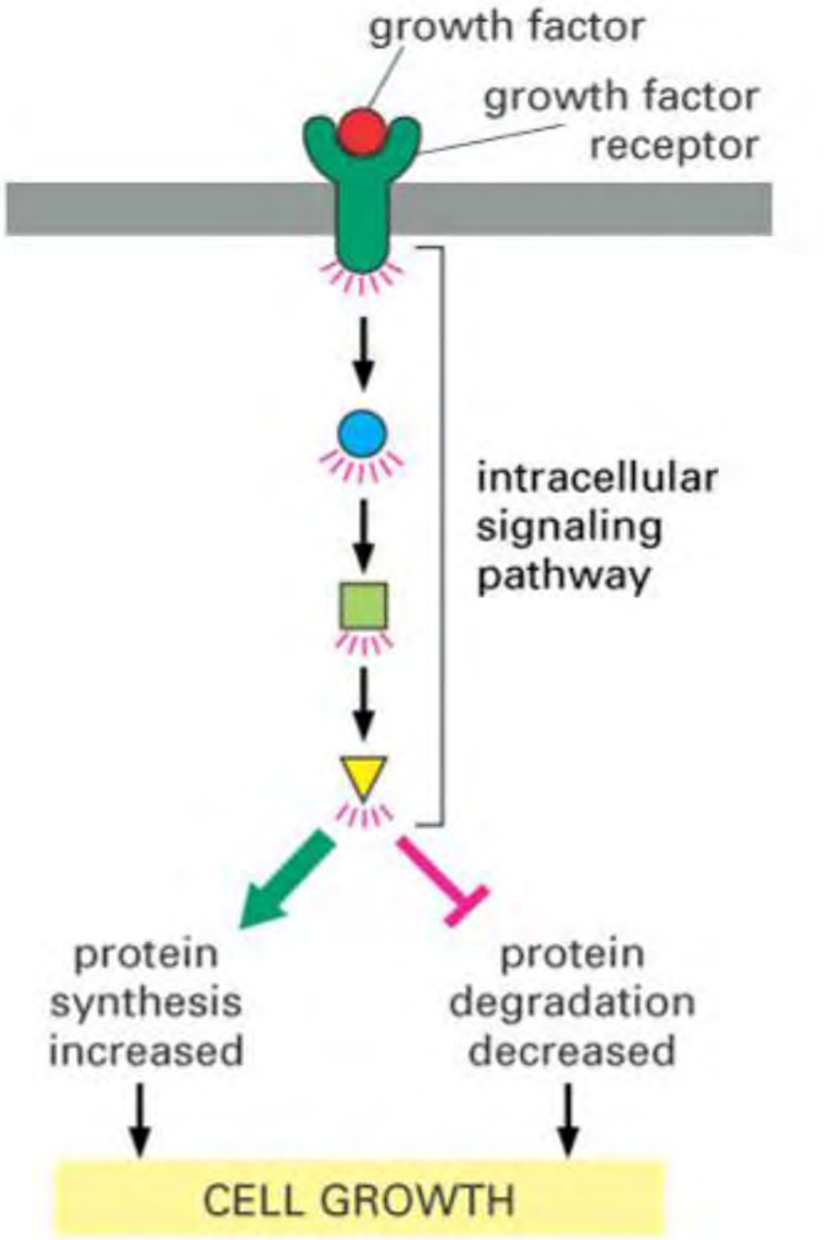
Pathguide: 547 resources http://www.pathguide.org/

STOP 1:
Pathway Profiling / Compound Mapping

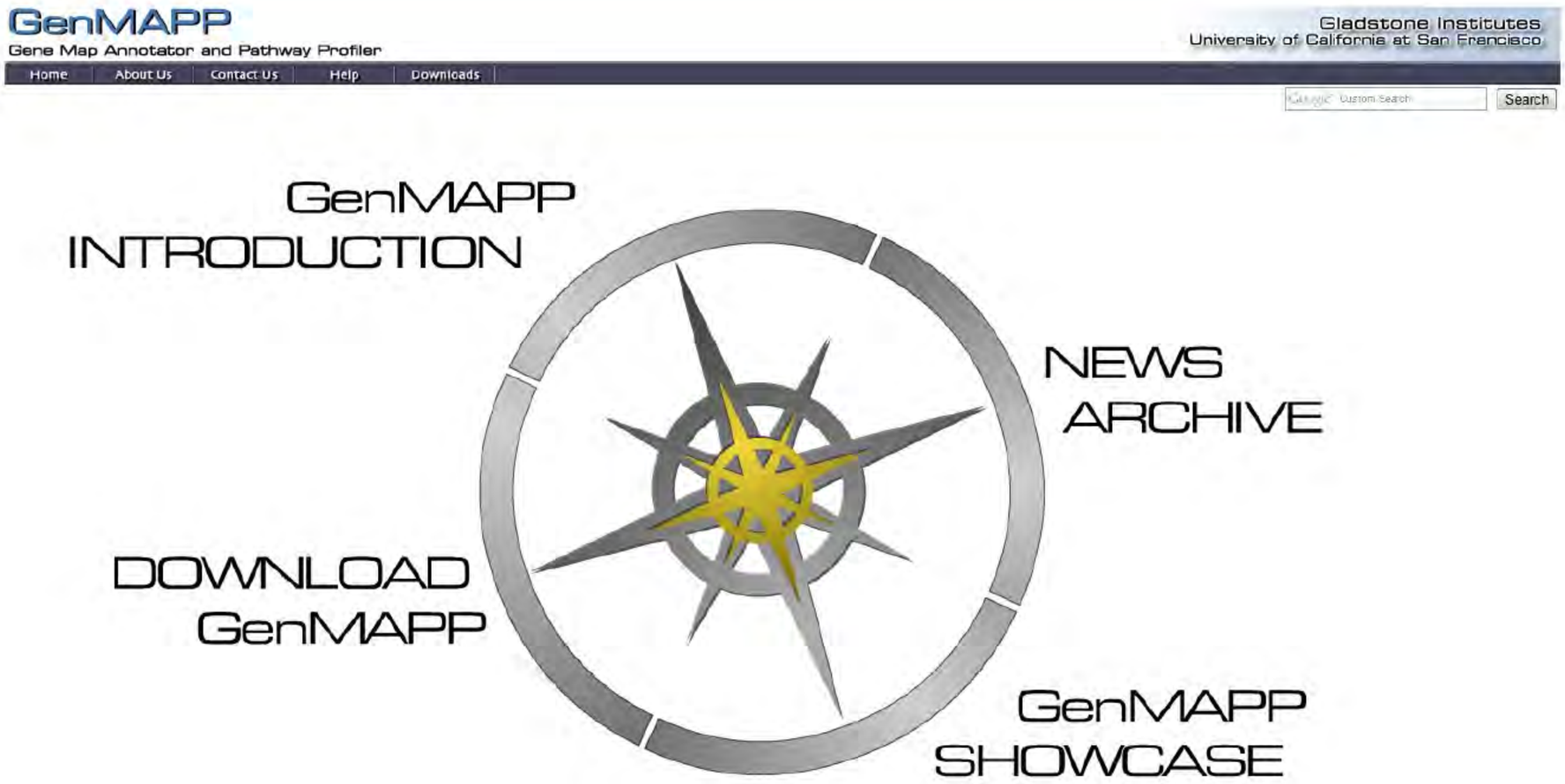
A representation of a peak (biochemical data) being associated with a compound in a database, then with a compound in a biological pathway map



Biological Pathways/ Signaling Pathways



GenMAPP 2.1

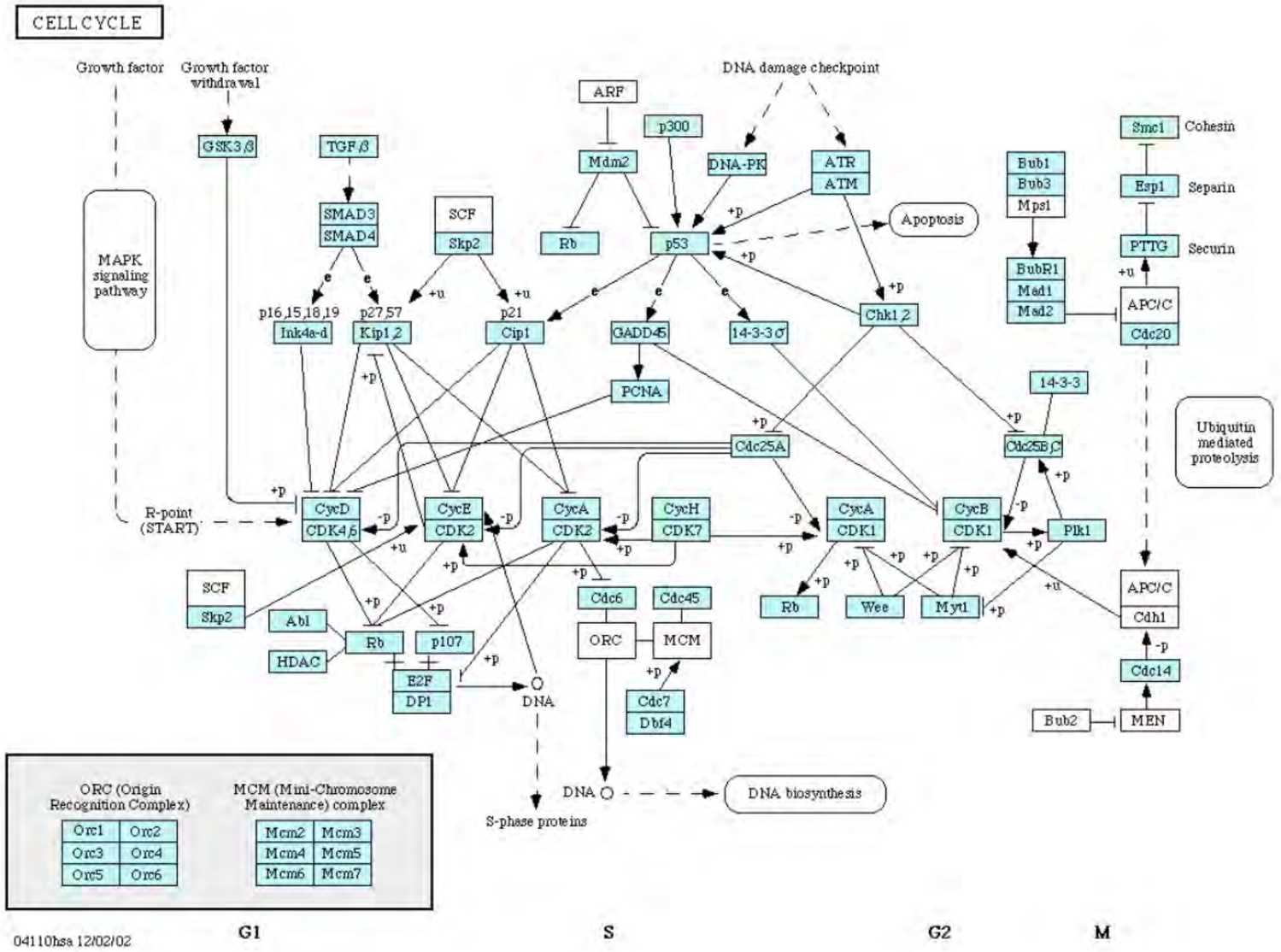


http://www.genmapp.org/concept.html

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/wajant530.swf

http://stke.sciencemag.org/content/sigtrans/suppl/2002/10/22/CMP_7966.DC1/wajant530.swf

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

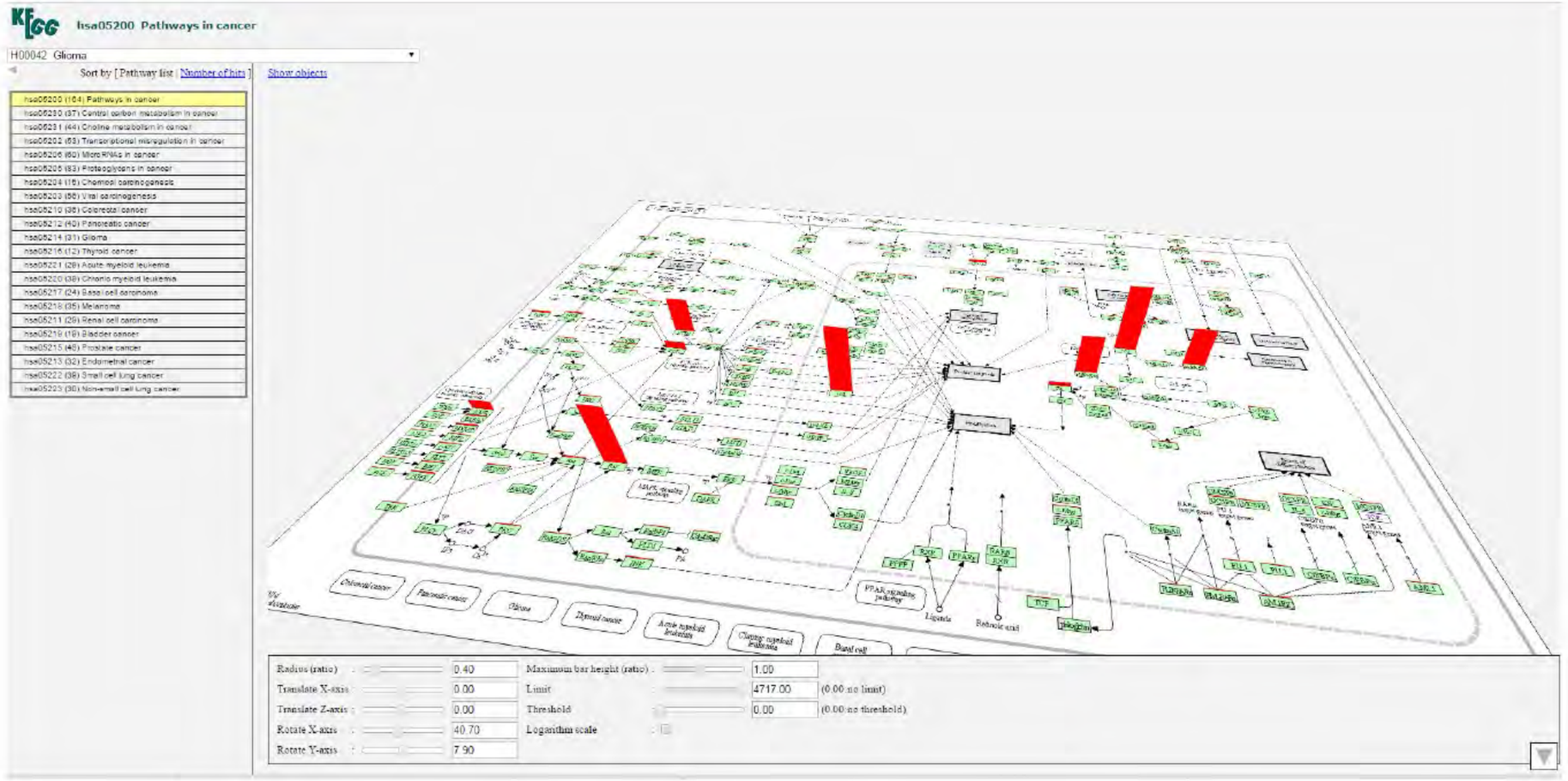
Current Statistics

KEGG Database as of 2015/1/23

Systems information		
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 information		
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 5,010,380,555
KEGG DGENES	Genes in draft genomes (26 eukaryotes)	654,883
KEGG MGENES	Genes in metagenomes (796 samples)	131,900,984
Chemical information		
KEGG COMPOUND	Metabolites and other small molecules	17,361
KEGG GLYCAN	Glycans	10,987
KEGG REACTION	Biochemical reactions	9,790
KEGG RPAIR	Reactant pair chemical transformations	14,874
KEGG RCLASS	Reaction class	2,949
KEGG ENZYME	Enzyme nomenclature	6,415
Health information		
KEGG DISEASE	Human diseases	1,402
KEGG DRUG	Drugs	10,142
KEGG DGROUP	Drug groups	1,741
KEGG ENVIRON	Crude drugs and health-related substances	850

http://www.kegg.jp/kegg/

KEGG Atlas: Cancer Pathways



Example: hsa05200 Pathways in cancer

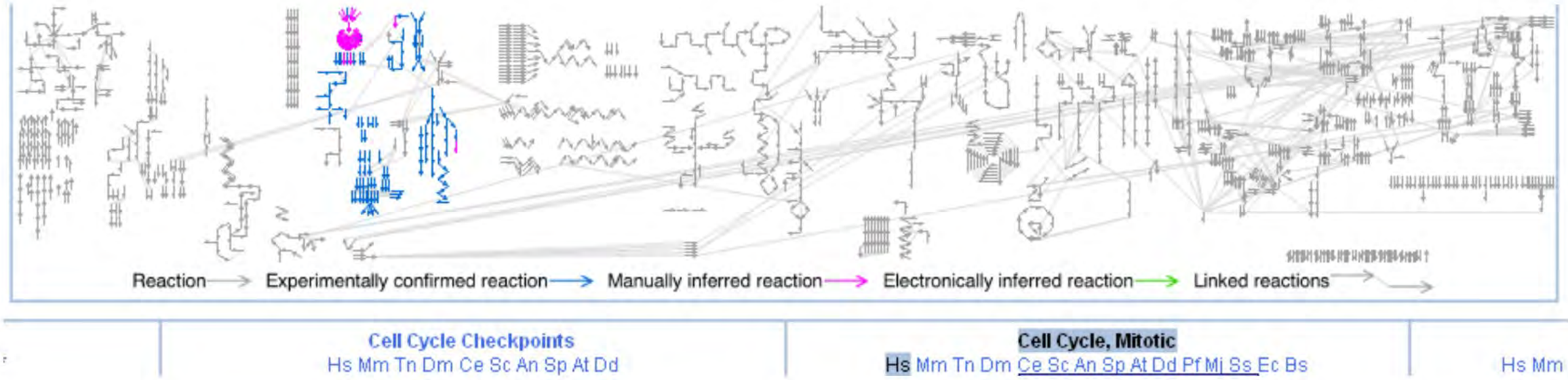
Reactome - a knowledgebase of biological processes

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

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



PathCards ALPHA

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

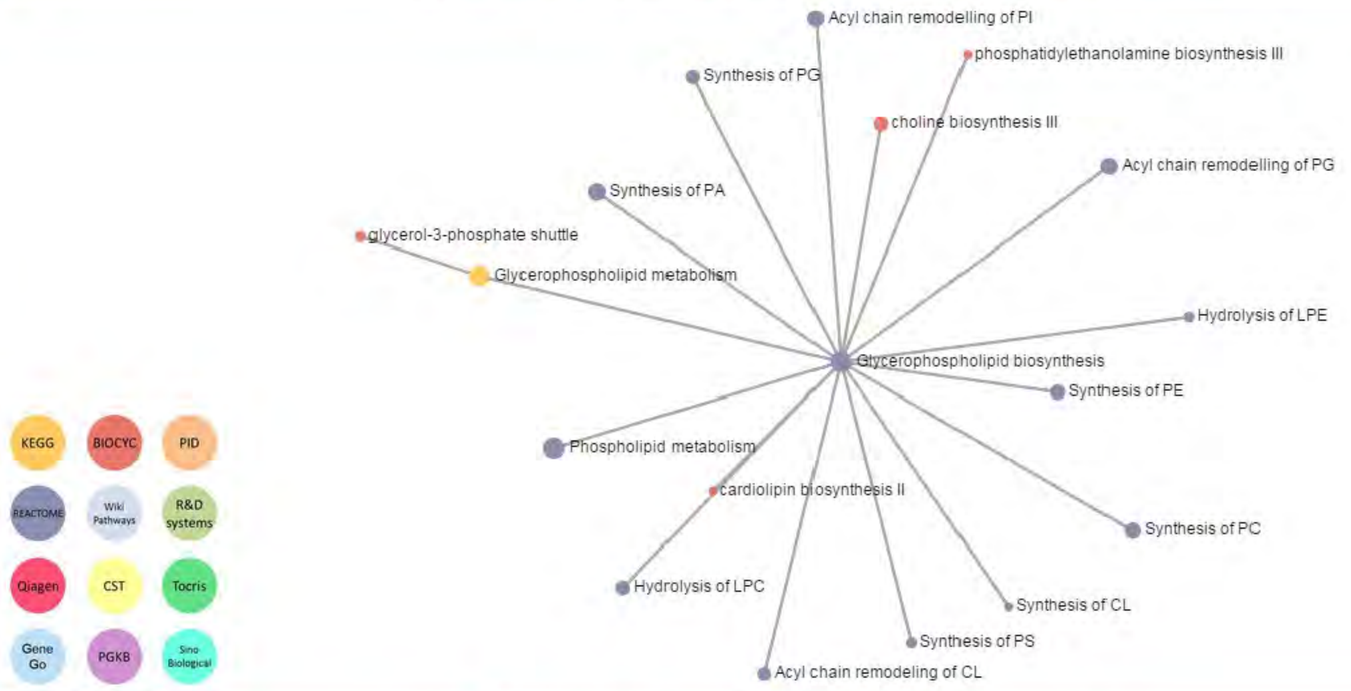
LifeMap SCIENCES

Welcome Guest
PathCards is part of GeneCards Plus | Sign in

Enter gene symbol, or a pathway name

Glycerophospholipid biosynthesis SuperPath

Pathway network for *Glycerophospholipid biosynthesis* SuperPath



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

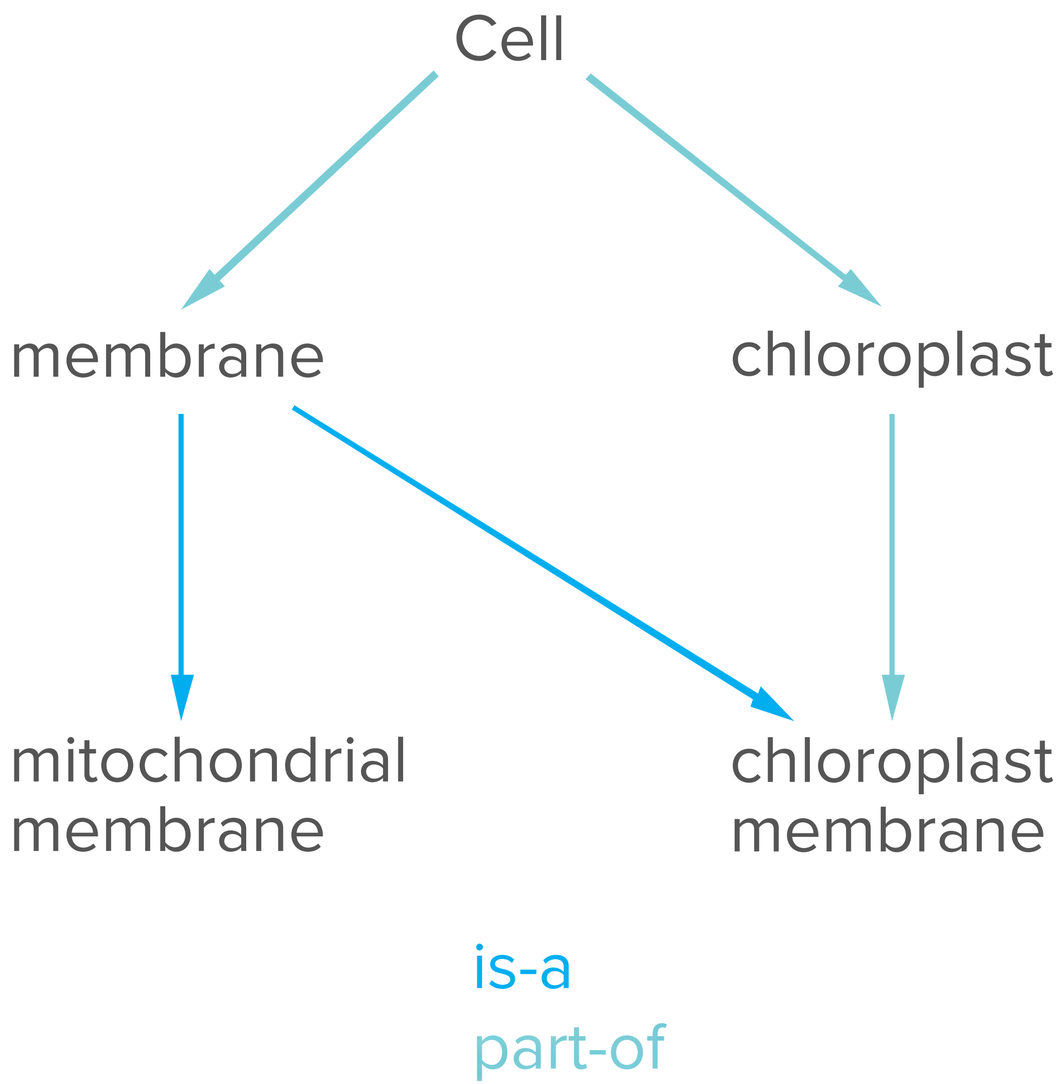
Biological Process

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)		Process (why)
Drive nail (into wood)		Carpentry
Drive stake (into soil)		Gardening
Smash roach		Pest Control
Clown’a juggling object		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)

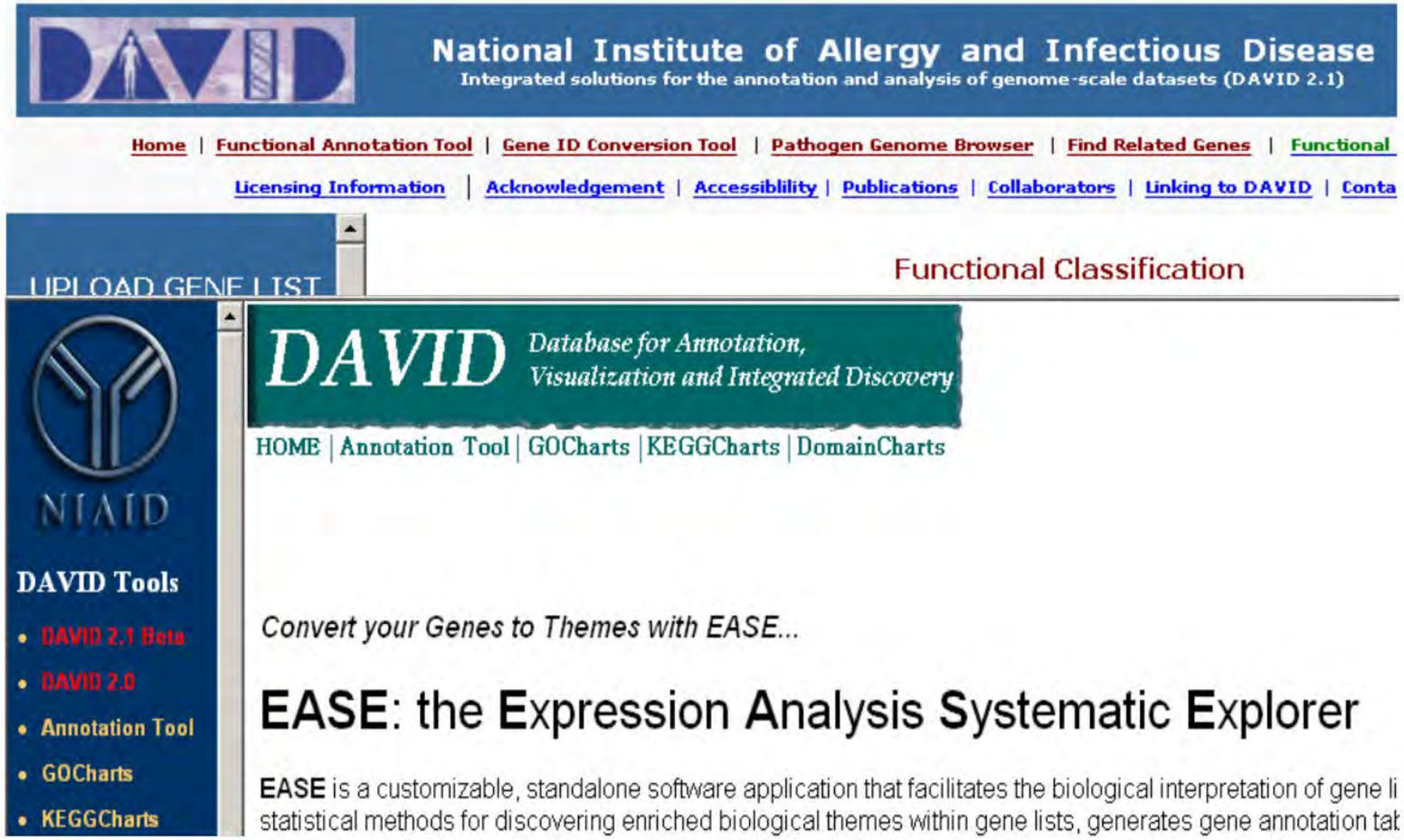
GO-based Tools

Functional profiling of microarray and proteomics data

Tools for gene expression/microarray analysis

BiNGO
CLENCH
DAVID
EASE
eGOn v2.0
ermineJ
FatiGO
FuncAssociate
FuncExpression
GARBAN
GeneMerge
GFINDer: Genome Function
GOArray
GOdist
GOODIES
GoMiner and MatchMiner
GOstat
GoSurfer
GO Term Finder
GOTM (Gene Ontology Tree Machine)

Combining Pathways and GO ontology enrichment analysis



DAVID

Database for Annotation, Visualisation and Integrated Discovery

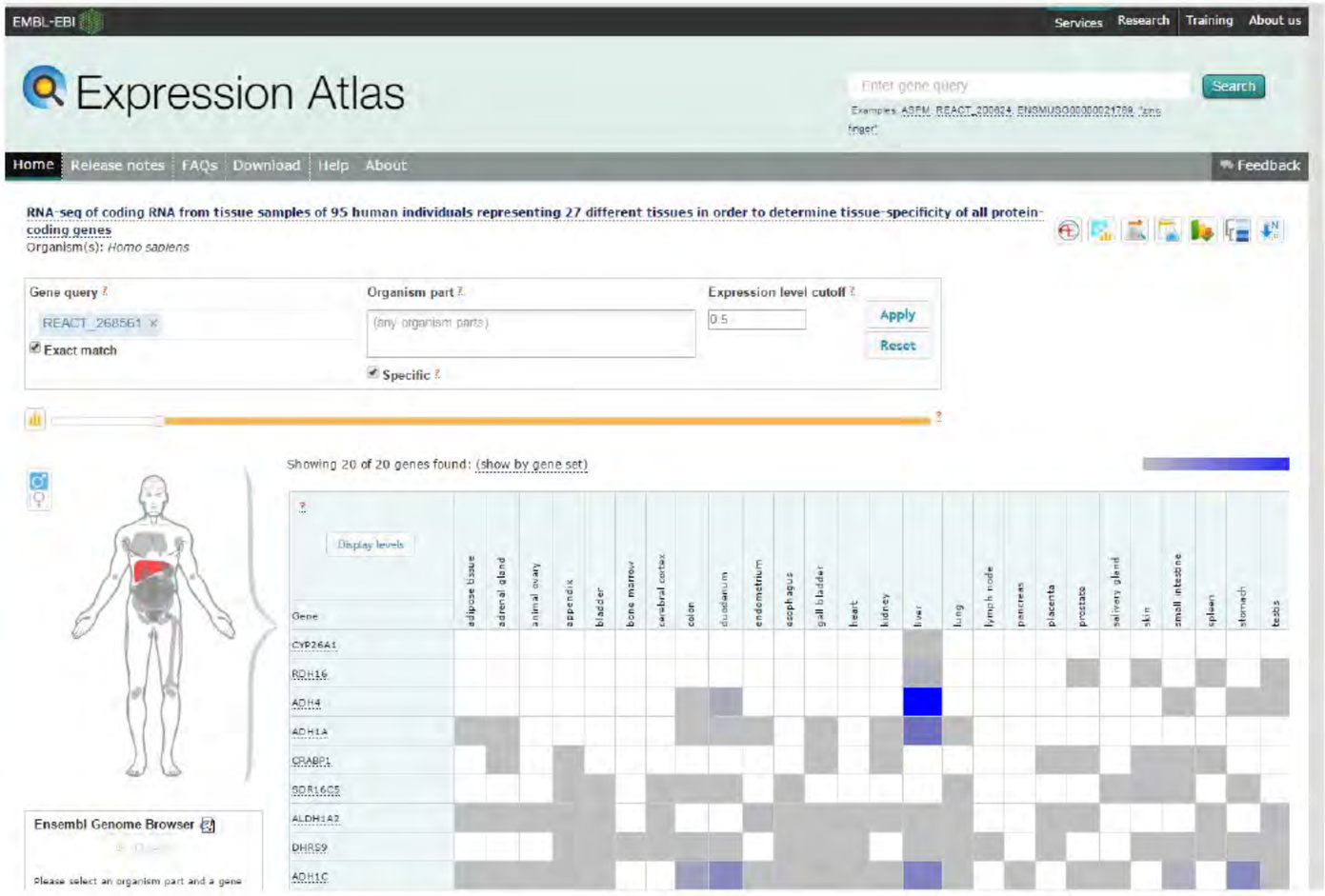
NEXT STOP: Next Generation of Systems Biology Tools

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?geneQuery=REACT_268561&serializedFilterFactors



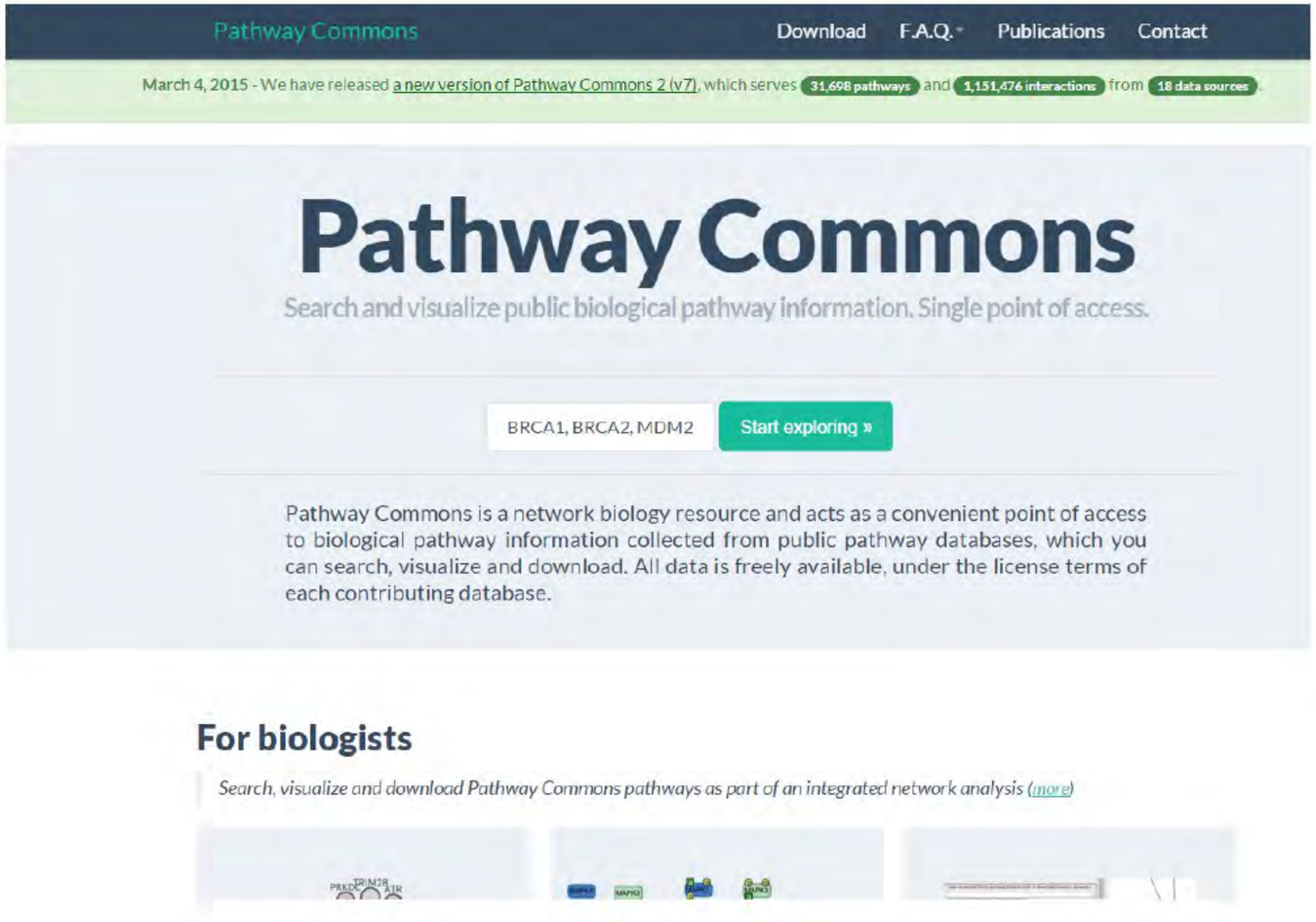
Pathway and Network analysis on-line: Public Domain

http://www.pathwaycommons.org/about/

http://www.ndexbio.org/

http://amp.pharm.mssm.edu/Enrichr/

CellMiner: http://discover.nci.nih.gov/cellminer/



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

GEO2Enrichr

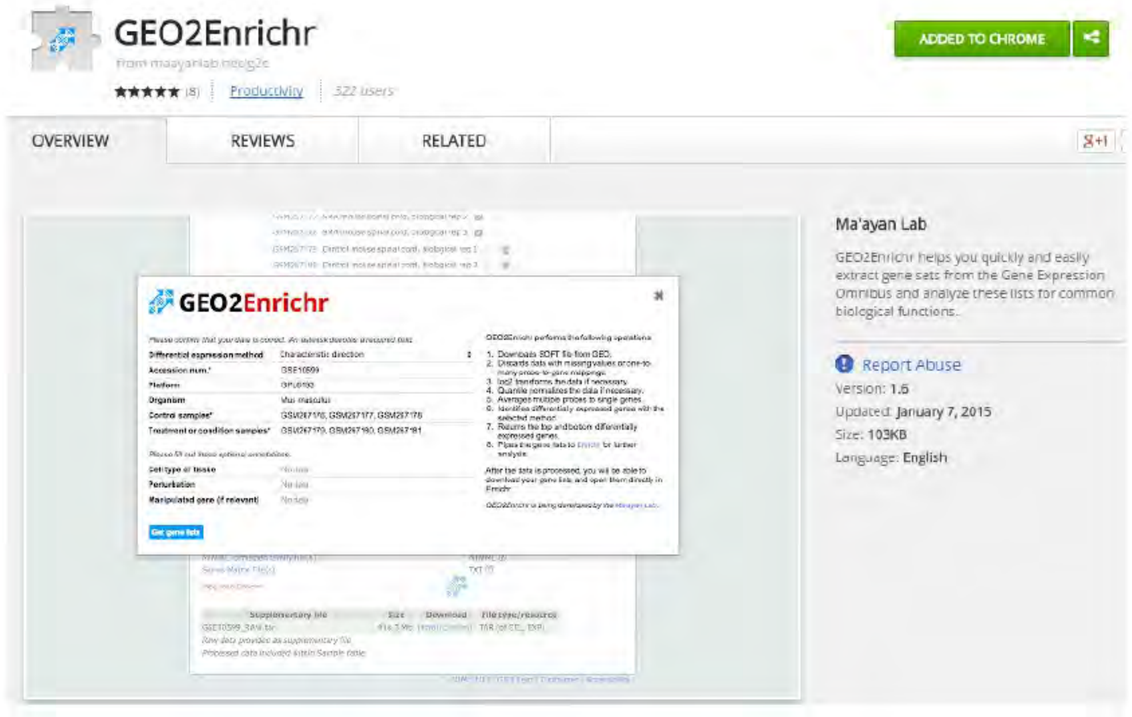
Google Chrome Extension to extract gene sets from GEO and analyze using Systems Biology Tools on-line

Data Public Repositories

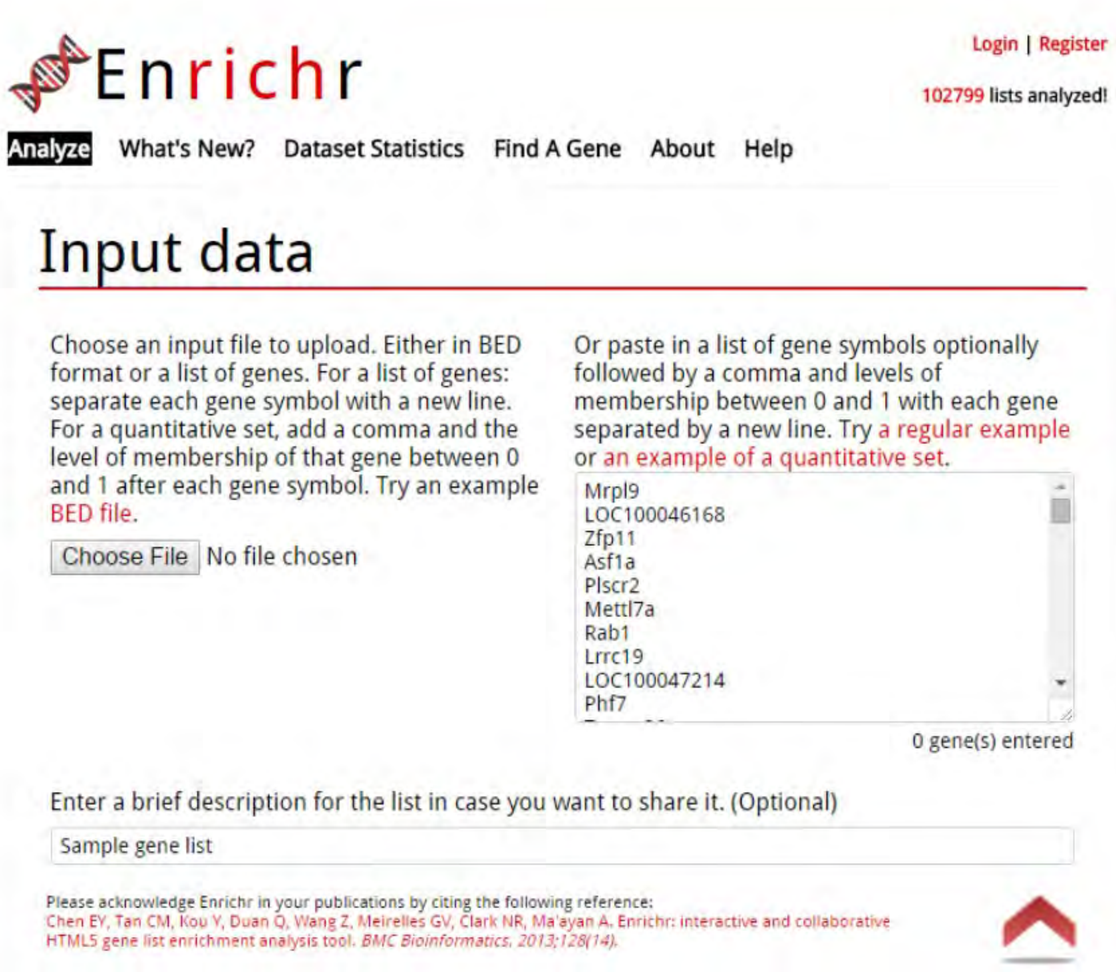
Gene Omnibus - GEO
Express Array
Short Reads Archive - SRA

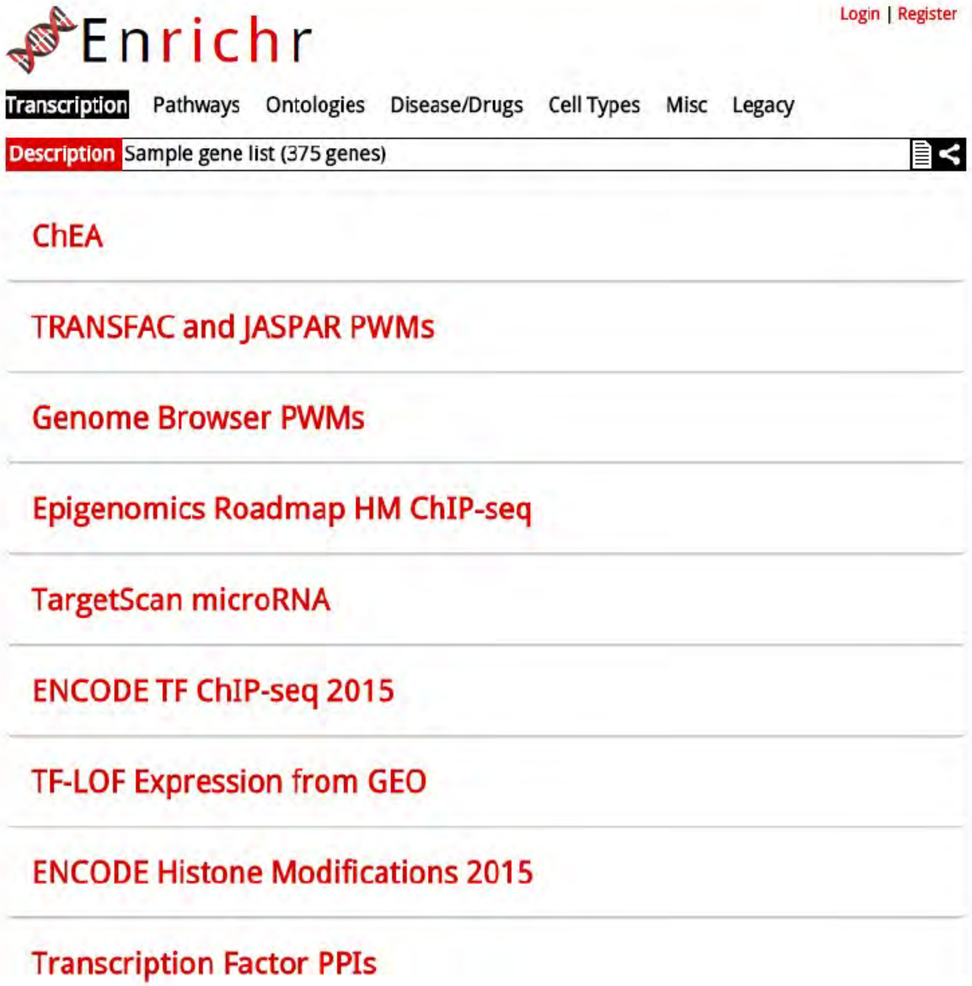
Systems Biology Tools

Pathway Analysis
Gene Ontology
Network Analysis

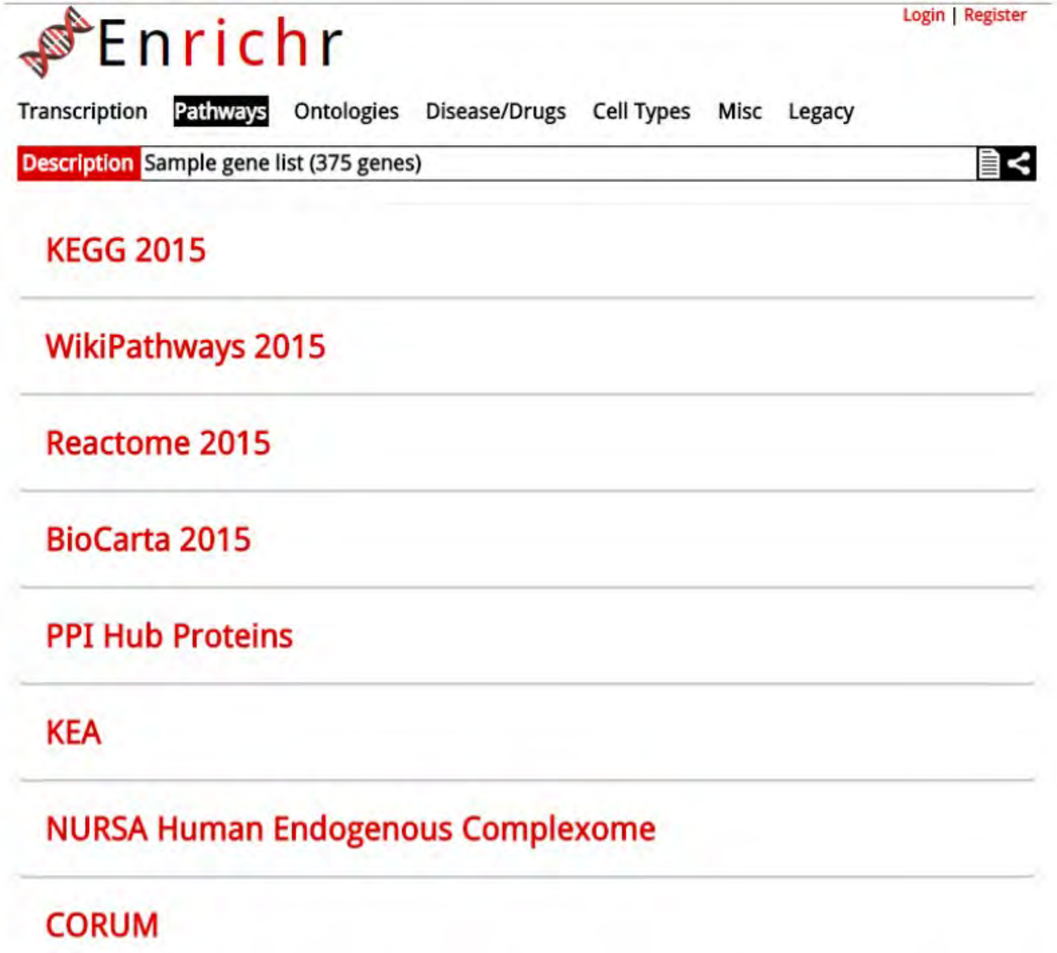


Enrichr


Enrichr
Login | Register
102799 lists analyzed!
Analyze
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
Login | Register
Transcription
Pathways
Ontologies
Disease/Drugs
Cell Types
Misc
Legacy
Description
Sample gene list (375 genes)
ChEA
TRANSFAC and JASPAR PWMs
Genome Browser PWMs
Epigenomics Roadmap HM ChIP-seq
TargetScan microRNA
ENCODE TF ChIP-seq 2015
TF-LOF Expression from GEO
ENCODE Histone Modifications 2015
Transcription Factor PPIs

Enrichr



Panther

Bar Graph | Table

Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Z-score	Combined Score
1	Beta1 adrenergic receptor signaling pathway*	0.02936	-1.80	1.29
2	Purine metabolism*	0.01387	-1.50	1.07
3	Metabotropic glutamate receptor group II pathway*	0.07184	-1.37	0.88
4	Thyrotropin-releasing hormone receptor signaling pathway*	0.1257	-1.12	0.72
5	Muscarinic acetylcholine receptor 2 and 4 signaling pathway*	0.1114	-1.09	0.70
6	Heme biosynthesis*	0.02116	-0.89	0.64
7	Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade*	0.07184	-0.91	0.58
8	Beta2 adrenergic receptor signaling pathway*	0.1352	1.61	0.00
9	Coenzyme A biosynthesis*	0.1352	3.69	0.00
10	Beta3 adrenergic receptor signaling pathway*	0.1601	2.51	0.00

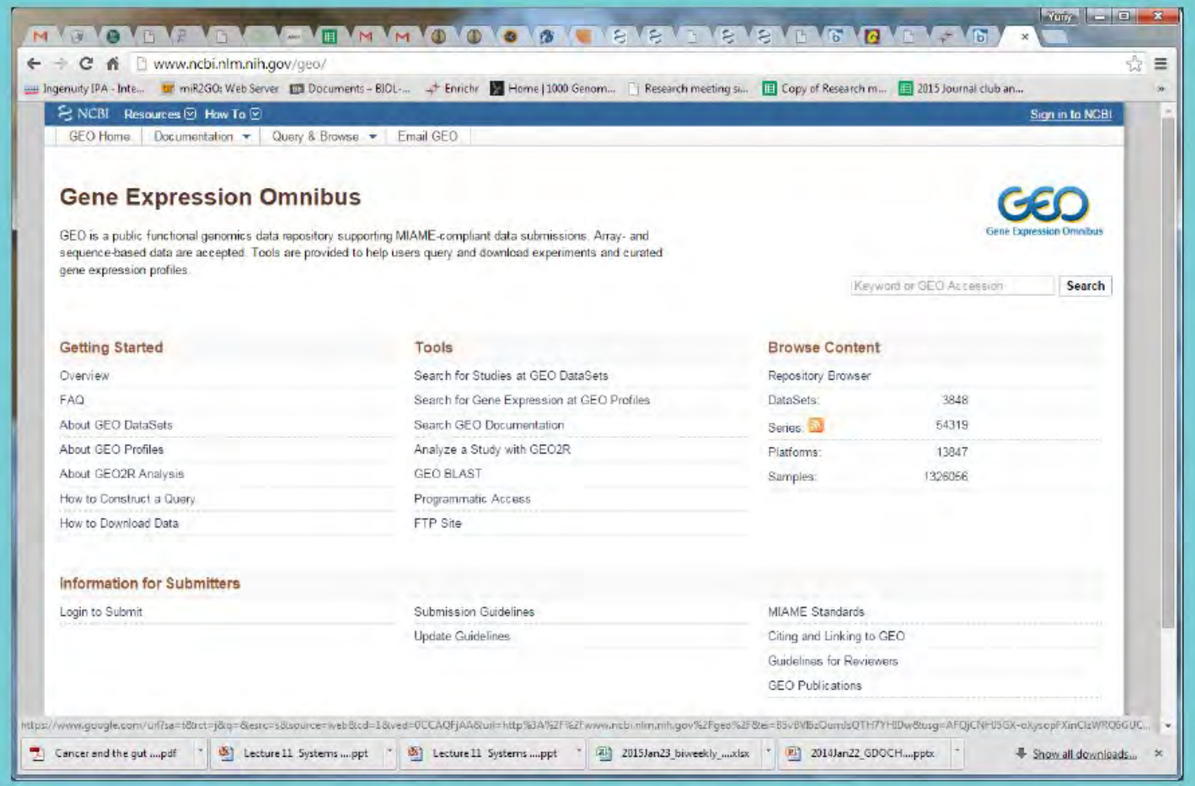
Showing 1 to 10 of 50 entries | Export entries to table

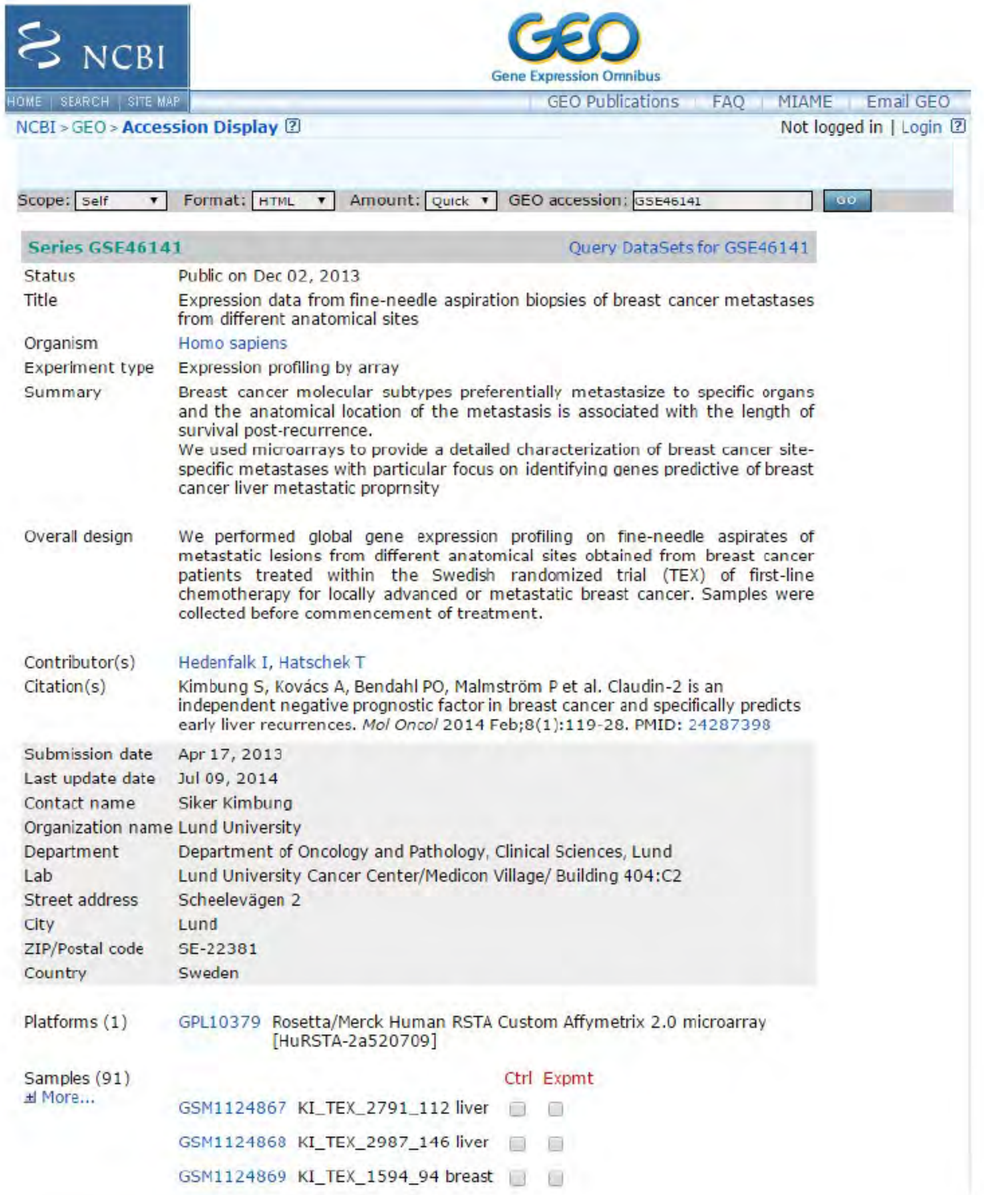
Previous Next

Terms marked with an * have an overlap of less than 5

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

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46141





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

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46141

Platforms (1)	GPL10379	Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray [HuRSTA-2a520709]

Samples (91) ⊞ More...

		Ctrl	Expmt
GSM1124867	KI_TEX_2791_112 liver	☐	☐
GSM1124868	KI_TEX_2987_146 liver	☐	☐
GSM1124869	KI_TEX_1594_94 breast	☐	☐

Relations

BioProject PRJNA197346

Analyze with GEO2R

Download family	Format
SOFT formatted family file(s)	SOFT ?
MINiML formatted family file(s)	MINiML ?
Series Matrix File(s)	TXT ?

Pipe into Enrichr

Supplementary file	Size	Download	File type/resource
GSE46141_Probeset_Annotation_20100204.txt.gz	9.3 Mb	(ftp)(http)	TXT
GSE46141_RAW.tar	469.9 Mb	(http)(custom)	TAR (of CEL)

Raw data provided as supplementary file

Processed data included within Sample table

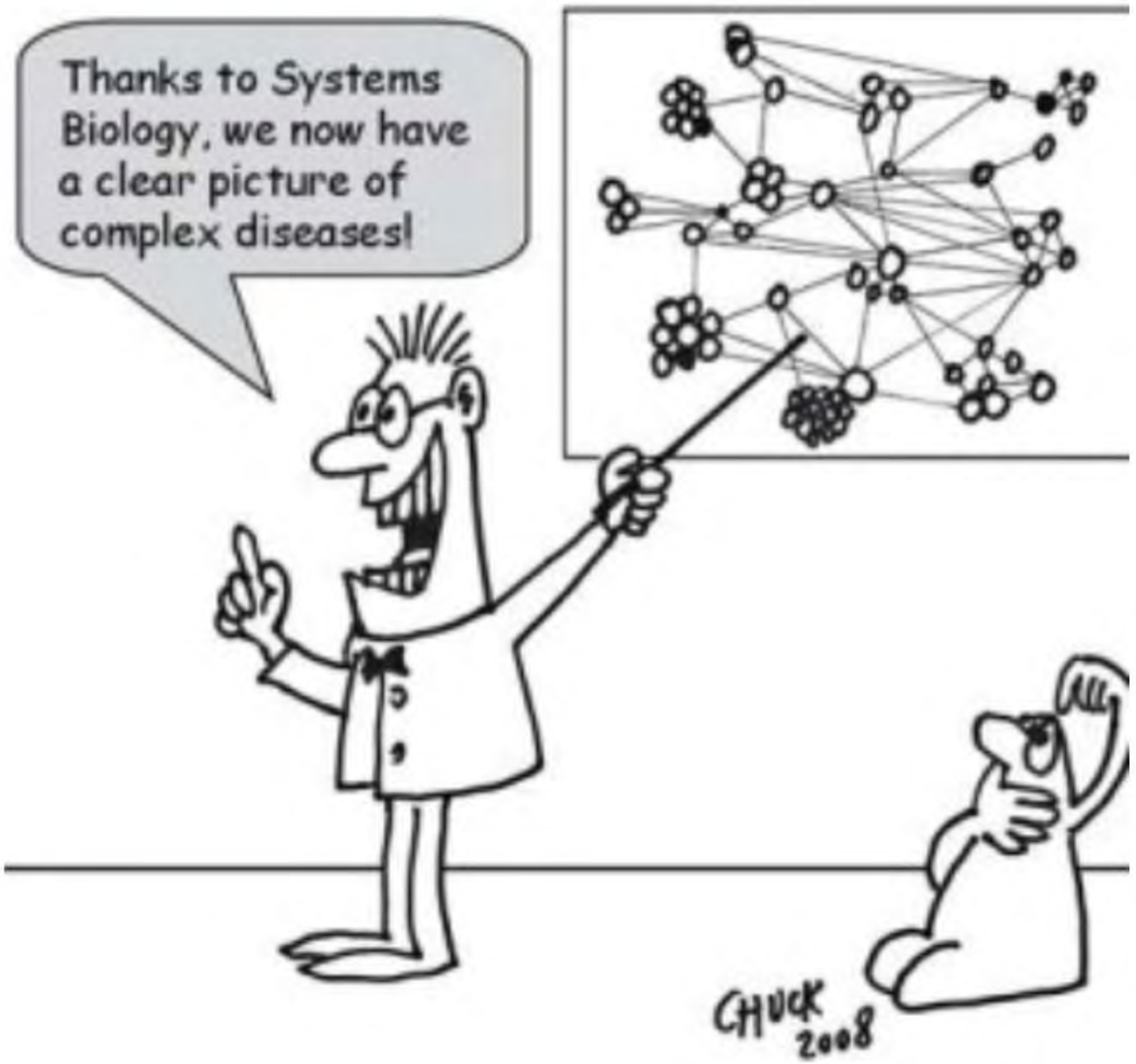
Contributor(s)	Hedenfalk I, Hatschek T
Citation(s)	Kimbung S, Kovács A, Bendahl PO, Malmström P et al. Claudin-2 is an independent negative prognostic factor in breast cancer and specifically predicts early liver recurrences. *Mol Oncol* 2014 Feb;8(1):119-28. PMID: 24287398
Submission date	Apr 17, 2013
Last update date	Jul 09, 2014
Contact name	Siker Kimbung
Organization name	Lund University
Department	Department of Oncology and Pathology, Clinical Sciences, Lund
Lab	Lund University Cancer Center/Medicon Village/ Building 404:C2
Street address	Scheelevägen 2
City	Lund
ZIP/Postal code	SE-22381
Country	Sweden

Platforms (1)	GPL10379	Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray [HuRSTA-2a520709]

Samples (91) ⊟ Less...

		Ctrl	Expmt
GSM1124867	KI_TEX_2791_112 liver	☐	☑
GSM1124868	KI_TEX_2987_146 liver	☐	☑
GSM1124869	KI_TEX_1594_94 breast	☑	☐
GSM1124870	KI_TEX_1583_8 breast	☑	☐
GSM1124871	KI_TEX_1524_35 lymph node	☐	☐
GSM1124872	KI_TEX_1467_82 skin local-regional	☐	☐
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	☐	☐

Questions?


Thanks to Systems Biology, we now have a clear picture of complex diseases!
CHUCK 2008