Gene List Enrichment Analysis - Statistics, Tools, Data Integration and Visualization

Aik Choon Tan, Ph.D.
Associate Professor of Bioinformatics
Division of Medical Oncology
Department of Medicine
aikchoon.tan@ucdenver.edu
10/9/2018
http://tanlab.ucdenver.edu/labHomePage/teaching/CANB7640/

Outline

- Why do enrichment analysis?
- Types of Enrichment analysis
- Statistics
- Annotation sources / databases and Tools
- Data Integration and Visualization tools

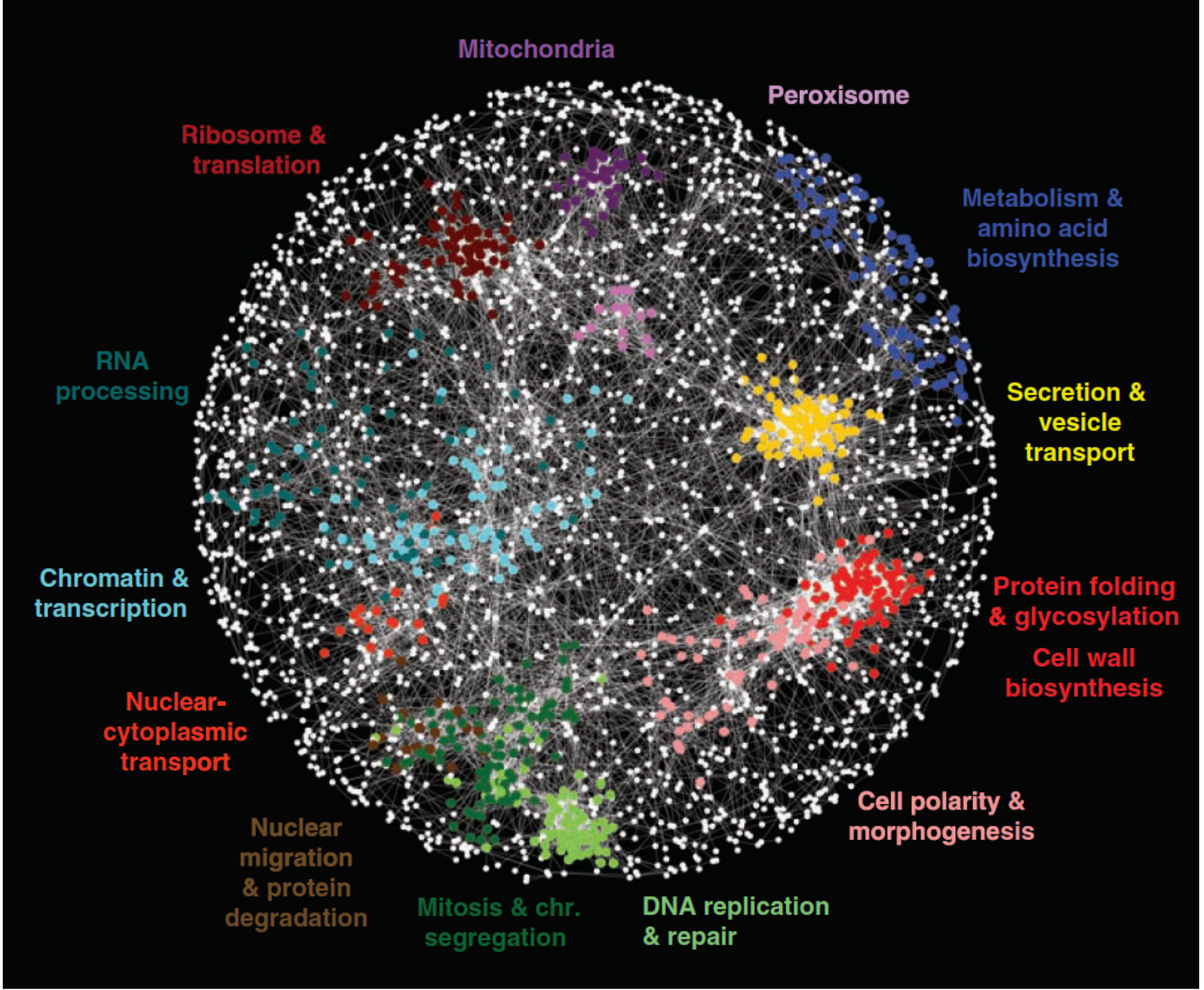
The Genetic Landscape of a Cell

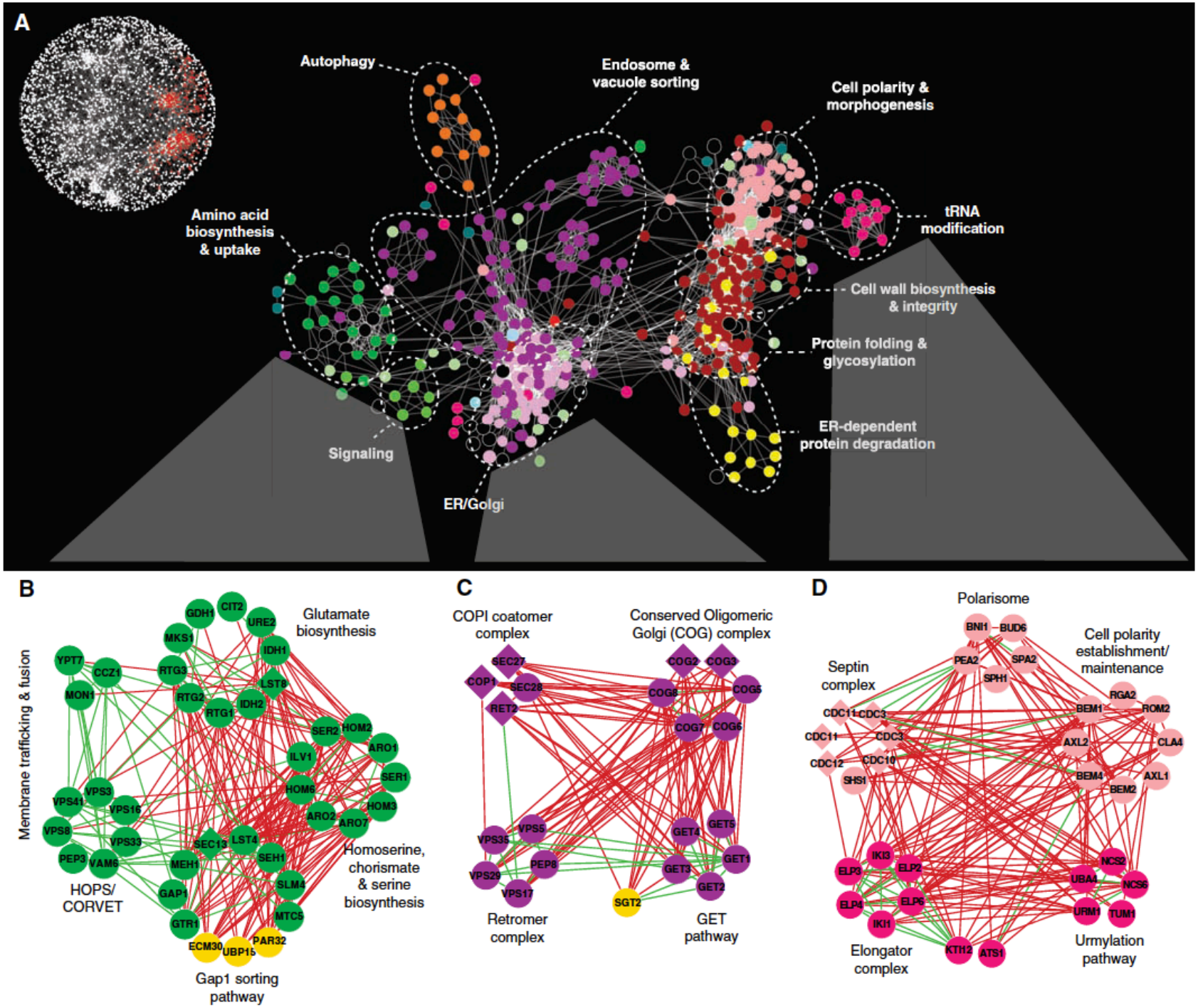
Michael Costanzo,[1,2]* Anastasia Baryshnikova,[1,2]* Jeremy Bellay,[3] Yungil Kim,[3] Eric D. Spear,[4] Carolyn S. Sevier,[4] Huiming Ding,[1,2] Judice L.Y. Koh,[1,2] Kiana Toufighi,[1,2] Sara Mostafavi,[1,5] Jeany Prinz,[1,2] Robert P. St. Onge,[6] Benjamin VanderSluis,[3] Taras Makhnevych,[7] Franco J. Vizeacoumar,[1,2] Solmaz Alizadeh,[1,2] Sondra Bahr,[1,2] Renee L. Brost,[1,2] Yiqun Chen,[1,2] Murat Cokol,[8] Raamesh Deshpande,[3] Zhijian Li,[1,2] Zhen-Yuan Lin,[9] Wendy Liang,[1,2] Michaela Marback,[1,2] Jadine Paw,[1,2] Bryan-Joseph San Luis,[1,2] Ermira Shuteriqi,[1,2] Amy Hin Yan Tong,[1,2] Nydia van Dyk,[1,2] Iain M. Wallace,[1,2,10] Joseph A. Whitney,[1,5] Matthew T. Weirauch,[11] Guoqing Zhong,[1,2] Hongwei Zhu,[1,2] Walid A. Houry,[7] Michael Brudno,[1,5] Sasan Ragibizadeh,[12] Balázs Papp,[13] Csaba Pál,[13] Frederick P. Roth,[8] Guri Giaever,[2,10] Corey Nislow,[1,2] Olga G. Troyanskaya,[14] Howard Bussey,[15] Gary D. Bader,[1,2] Anne-Claude Gingras,[9] Quaid D. Morris,[1,2,5] Philip M. Kim,[1,2] Chris A. Kaiser,[4] Chad L. Myers,[3]† Brenda J. Andrews,[1,2]† Charles Boone[1,2]†

A genome-scale genetic interaction map was constructed by examining 5.4 million gene-gene pairs for synthetic genetic interactions, generating quantitative genetic interaction profiles for ~75% of all genes in the budding yeast, *Saccharomyces cerevisiae*. A network based on genetic interaction profiles reveals a functional map of the cell in which genes of similar biological processes cluster together in coherent subsets, and highly correlated profiles delineate specific pathways to define gene function. The global network identifies functional cross-connections between all bioprocesses, mapping a cellular wiring diagram of pleiotropy. Genetic interaction degree correlated with a number of different gene attributes, which may be informative about genetic network hubs in other organisms. We also demonstrate that extensive and unbiased mapping of the genetic landscape provides a key for interpretation of chemical-genetic interactions and drug target identification.

Systematic deletion analysis in the budding yeast, *Saccharomyces cerevisiae*, demonstrates that the majority of its ~6000 genes are individually dispensable, with only a relatively

[1]Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [2]Department of Molecular Genetics, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [3]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [5]Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada. [6]Department of Biochemistry, Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304, USA. [7]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [8]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [9]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. [10]Department of Pharmacy, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [11]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA. [12]S&P Robotics, Inc., 1181 Finch Avenue West, North York, Ontario M3J 2V8, Canada. [13]Institute of Biochemistry, Biological Research Center, H-6701 Szeged, Hungary. [14]Department of Computer Science, Lewis-Sigler Institute for Integrative Genomics, Carl Icahn Laboratory, Princeton University, Princeton, NJ 08544, USA. [15]Biology Department, McGill University, Montreal, Quebec H3A 1B1, Canada.

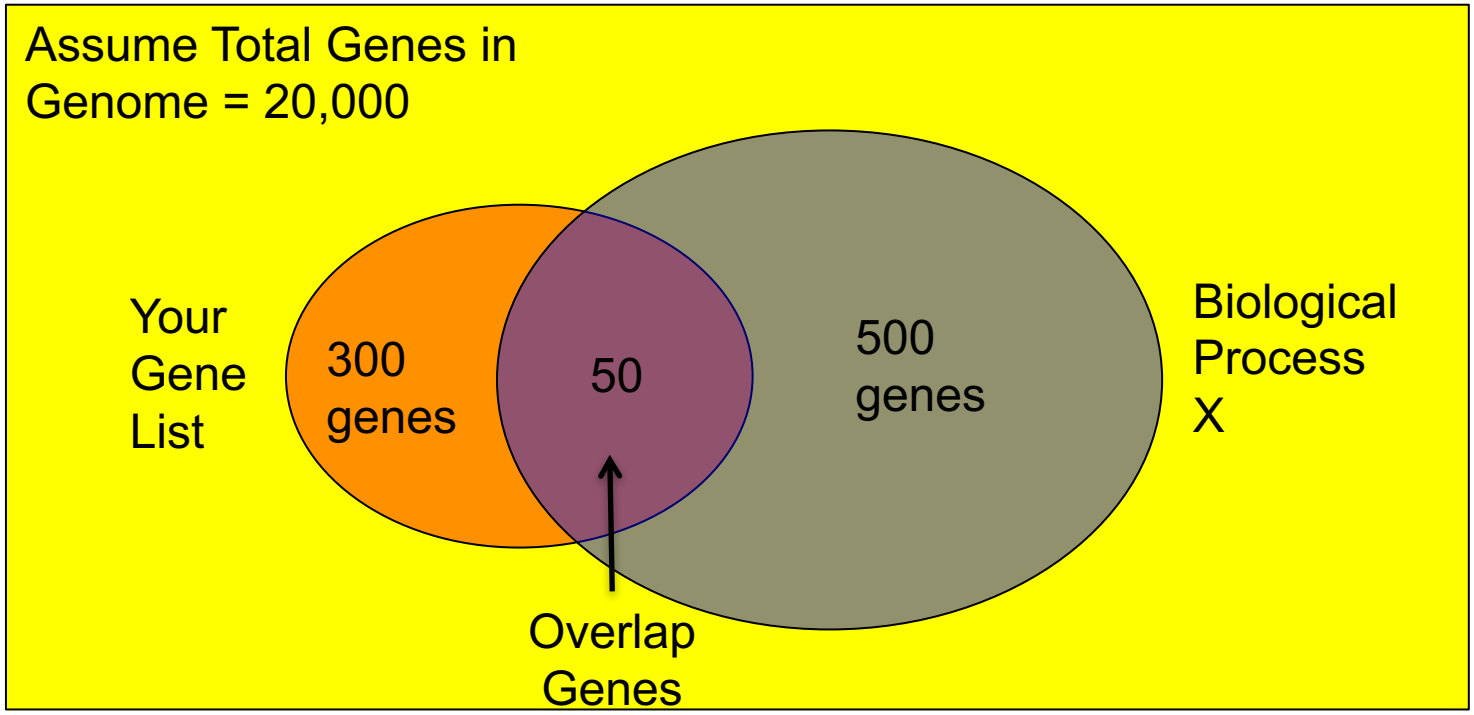

Mitochondria
Peroxisome
Ribosome & translation
Metabolism & amino acid biosynthesis
RNA processing
Secretion & vesicle transport
Chromatin & transcription
Protein folding & glycosylation
Cell wall biosynthesis
Nuclear-cytoplasmic transport
Cell polarity & morphogenesis
Nuclear migration & protein degradation
Mitosis & chr. segregation
DNA replication & repair


A
Autophagy
Endosome & vacuole sorting
Cell polarity & morphogenesis
Amino acid biosynthesis & uptake
tRNA modification
Cell wall biosynthesis & integrity
Protein folding & glycosylation
ER-dependent protein degradation
Signaling
ER/Golgi
B
Glutamate biosynthesis
Membrane trafficking & fusion
Homoserine, chorismate & serine biosynthesis
HOPS/ CORVET
Gap1 sorting pathway
GDH1
CIT2
URE2
MKS1
IDH1
YPT7
CCZ1
RTG3
LST8
MON1
RTG2
IDH2
RTG1
SER2
HOM2
ARO1
ILV1
SER1
VPS3
VPS41
HOM6
HOM3
VPS16
ARO2
ARO7
VPS8
VPS33
SEC13
LST4
PEP3
VAM6
MEH1
SEH1
SLM4
GAP1
MTC5
GTR1
ECM30
UBP15
PAR32
C
COPI coatomer complex
Conserved Oligomeric Golgi (COG) complex
Retromer complex
GET pathway
SEC27
COG2
COG3
COP1
SEC28
COG8
COG5
RET2
COG7
COG6
VPS5
GET5
VPS35
GET4
PEP8
GET1
GET3
VPS29
GET2
VPS17
SGT2
D
Polarisome
Septin complex
Cell polarity establishment/ maintenance
Elongator complex
Urmylation pathway
BNI1
BUD6
PEA2
SPA2
SPH1
RGA2
BEM1
ROM2
CDC11
CDC3
AXL2
CLA4
CDC12
CDC10
SHS1
BEM4
AXL1
BEM2
IKI3
NCS2
ELP3
ELP2
UBA4
NCS6
ELP4
ELP6
URM1
TUM1
IKI1
KTI12
ATS1

Why Do Enrichment Analysis?

- From “omics” data and analysis, generate list(s) of “interesting” gene lists (List A)
- Genes act in concert to drive biological processes (List B)
- Can we compare the two lists (Lists A and B)
- Is the overlap different than expected? (p-value)
- Is List A contains genes involved in List B biological process? (biology)

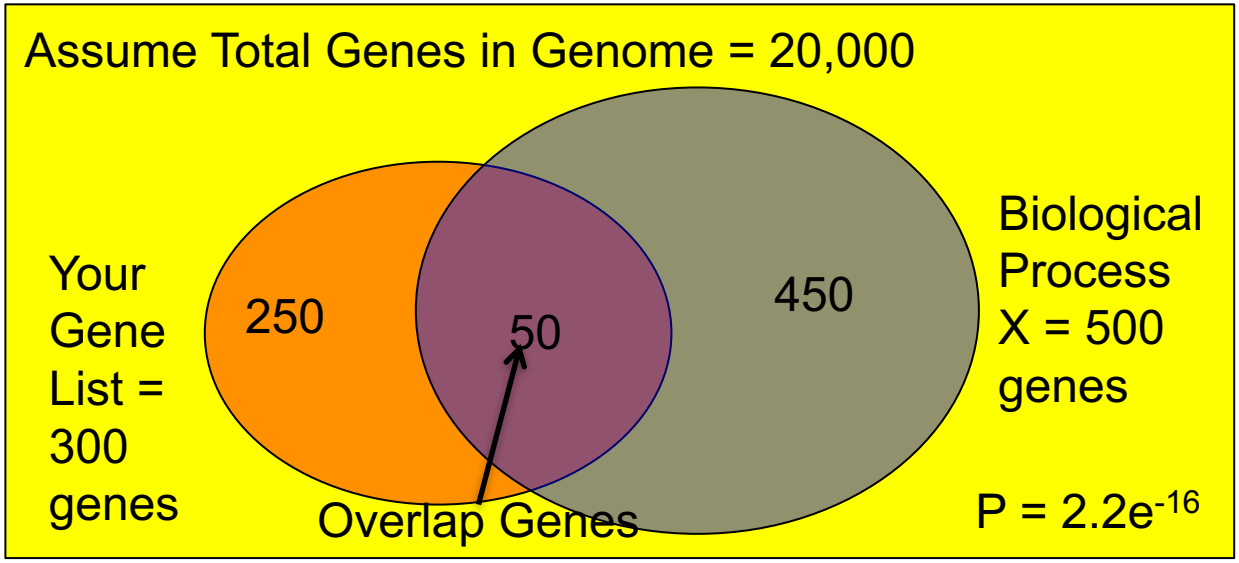
Idea

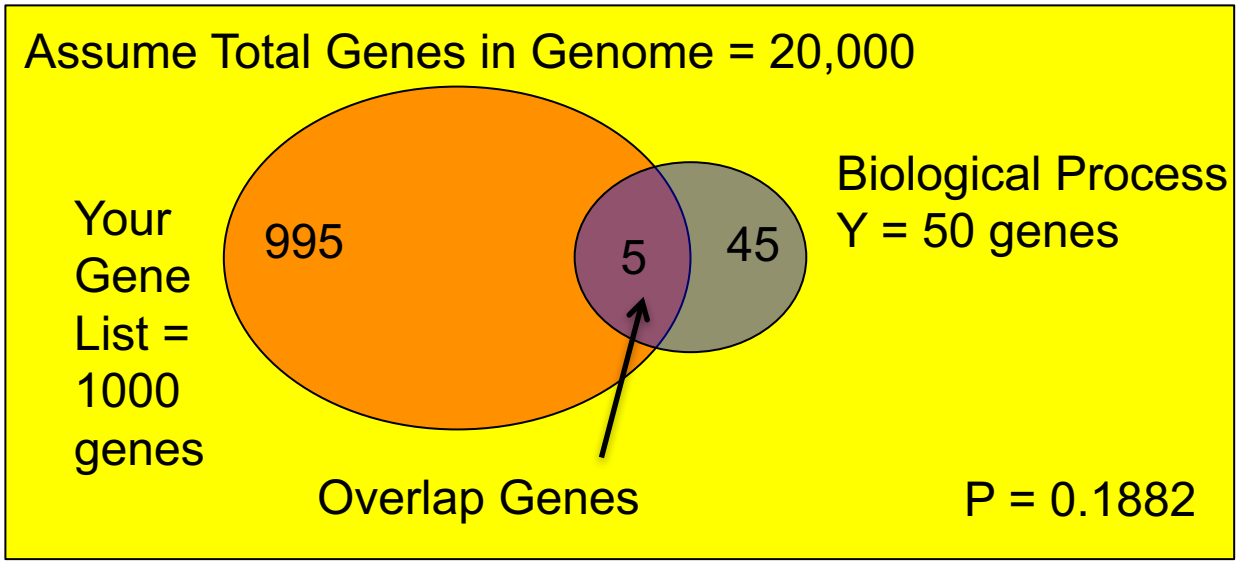

Assume Total Genes in Genome = 20,000
Your Gene List
300 genes
50
500 genes
Biological Process X
Overlap Genes

Different Tests for Enrichment

- Fisher's exact
- Chi-square
- Binomial
- Hypergeometric
- Kolmogorov-Smirnov *(This is implemented in GSEA)*
- Permutation
- …

Statistical Test for Enrichment





Statistics to Test for Enrichment

- What is the chance of observing enrichment at least this extreme due to chance?
- Different tests produce very different ranges of p-values
- All test for over-enrichment, some test for under-enrichment
- Recommendation: Use p-values as a tool to rank gene sets and don't take them literally (remember, biology trumps statistics)
- Useful to correct for multiple testing (e.g. FDR)

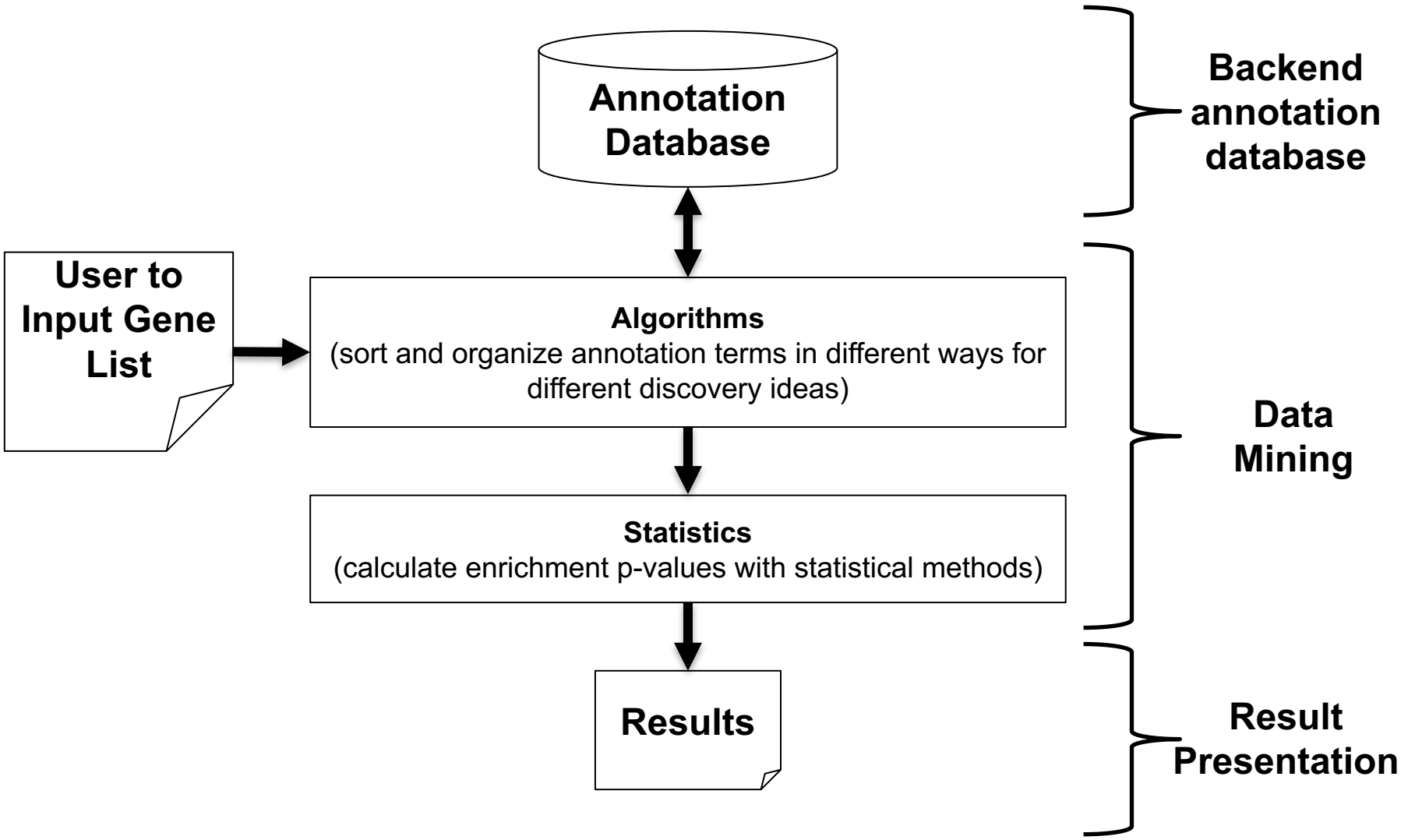
Input for Enrichment Test

- Background gene set
  - All genes that could appear in your gene list (usually, assume all the genes in the genome, e.g. 20,000 genes for human genome)
- Annotations / Gene sets
  - Which annotation sources / databases?
  - Some annotation terms may be subsets of other terms (e.g. in Gene Ontology)
- Goal: To identify gene sets/concepts with biological significance from your gene list

In Practice

- Choose a tool / program that
  - Includes your species
  - Includes your genes or probe identifiers (and do the matching from probes to genes)
  - Up-to-date annotation
  - Able to define background (if possible)
  - Able to select gene sets/annotations/concepts
- Try at least a few tools and compare
  - If the biology is real, you will find it from different tools
- Get recommendations from bioinformatics/statistics for interpreting the results

Gene List Enrichment Analysis



(Adapted from Huang et al NAR 2008)

So, which database(s) and tool(s) to use?

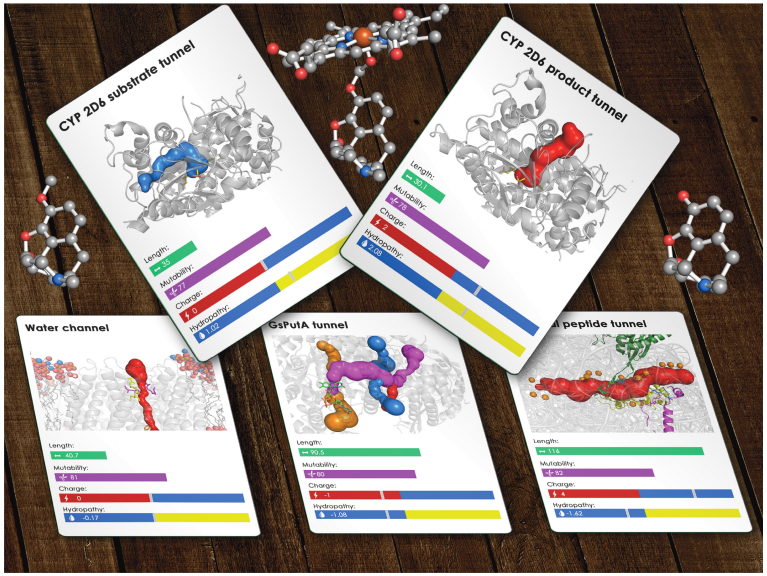
181 databases,
Online: 1737 databases
http://www.oxfordjournals.org/nar/database/c/

PRINT ISSN: 0305-1048
ONLINE ISSN: 1362-4962

DATABASE ISSUE (Jan)

Nucleic Acids Research

VOLUME 46 DATABASE ISSUE JANUARY 4, 2018
https://academic.oup.com/nar



OXFORD UNIVERSITY PRESS

Open Access
No barriers to access – all articles freely available online

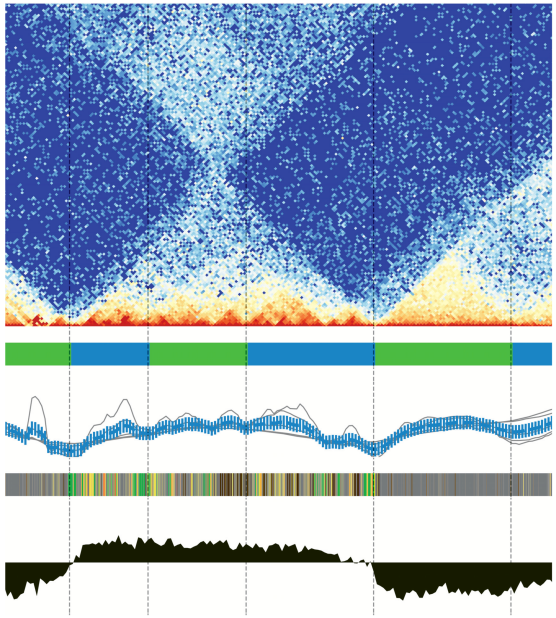
85 web servers

PRINT ISSN: 0305-1048
ONLINE ISSN: 1362-4962

WEB SERVER ISSUE (July)

Nucleic Acids Research

VOLUME 46 WEB SERVER ISSUE JULY 2, 2018
https://academic.oup.com/nar



OXFORD UNIVERSITY PRESS

Open Access
No barriers to access – all articles freely available online

Survey

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Review

Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges

Purvesh Khatri[1,2]*, Marina Sirota[1,2], Atul J. Butte[1,2]*

1 Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, United States of America, **2** Lucile Packard Children's Hospital, Palo Alto, California, United States of America

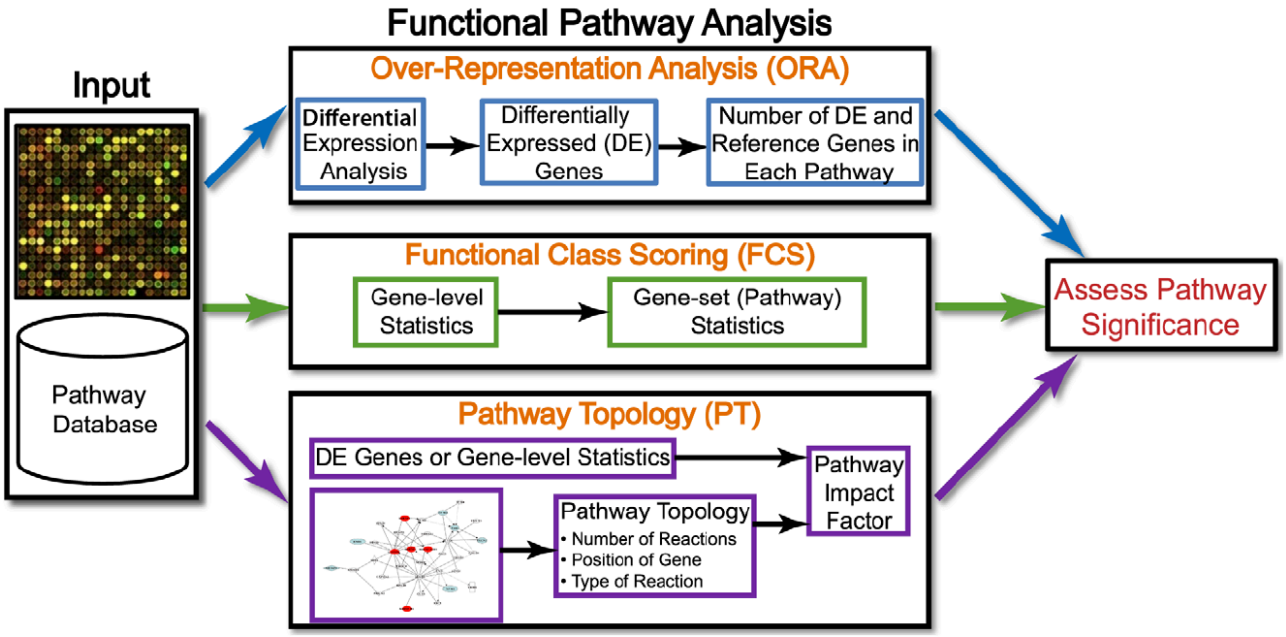


Table 1. Examples of pathway analysis tools in each generation.

Name	Availability	Reference
ORA tools		
Onto-Express	Web (http://vortex.cs.wayne.edu)	[4,5]
GenMAPP	Standalone (http://www.genmapp.org)	[11,71]
GoMiner	Standalone, Web (http://discover.nci.nih.gov/gominer)	[72,73]
FatiGO	Web (http://babelomics.bioinfo.cipf.es)	[74]
GOstat	Web (http://gostat.wehi.edu.au)	[7]
FuncAssociate	Web (http://llama.mshri.on.ca/funcassociate/)	[6]
GOToolBox	Web (http://genome.crg.es/GOToolBox/)	[10]
GeneMerge	Standalone, Web (http://genemerge.cbcb.umd.edu/)	[9]
GOEAST	Web (http://omicslab.genetics.ac.cn/GOEAST/)	[75]
ClueGO	Standalone (http://www.ici.upmc.fr/cluego/)	[76]
FunSpec	Web (http://funspec.med.utoronto.ca/)	[77]
GARBAN	Web	[78]
GO::TermFinder	Standalone (http://search.cpan.org/dist/GO-TermFinder/)	[8]
WebGestalt	Web (http://bioinfo.vanderbilt.edu/webgestalt/)	[79]
agriGO	Web (http://bioinfo.cau.edu.cn/agriGO/)	[80]
GOFFA	Standalone, Web (http://edkb.fda.gov/webstart/arraytrack/)	[81]
WEGO	Web (http://wego.genomics.org.cn/cgi-bin/wego/index.pl)	[82]
FCS tools		
GSEA	Standalone (http://www.broadinstitute.org/gsea/)	[21,29]
sigPathway	Standalone (BioConductor)	[22]
Category	Standalone (BioConductor)	[24]
SAFE	Standalone (BioConductor)	[30]
GlobalTest	Standalone (BioConductor)	[15]
PCOT2	Standalone (BioConductor)	[17]
SAM-GS	Standalone (http://www.ualberta.ca/~yyasui/software.html)	[83]
Catmap	Standalone (http://bioinfo.thep.lu.se/catmap.html)	[84]
T-profiler	Web (http://www.t-profiler.org)	[85]
FunCluster	Standalone (http://corneliu.henegar.info/FunCluster.htm)	[86]
GeneTrail	Web (http://genetrail.bioinf.uni-sb.de)	[87]
GAzer	Web	[88]
PT-based tools		
ScorePAGE	No implementation available	[37]
Pathway-Express	Web (http://vortex.cs.wayne.edu)	[38,39]
SPIA	Standalone (BioConductor)	[40]
NetGSA	No implementation available	[43]

doi:10.1371/journal.pcbi.1002375.t001

Survey

Published online 25 November 2008

Nucleic Acids Research, 2009, Vol. 37, No. 1 ***1–13***
doi:10.1093/nar/gkn923

SURVEY AND SUMMARY

Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists

Da Wei Huang, Brad T. Sherman and Richard A. Lempicki*

Laboratory of Immunopathogenesis and Bioinformatics, Clinical Services Program, SAIC-Frederick, Inc., National Cancer Institute at Frederick, Frederick, MD 21702, USA

Class I – Singular enrichment analysis (SEA) = ORA
Class II – Gene set enrichment analysis (GSEA) = FCS
Class III – Modular enrichment analysis (MEA) = PT

Enrichment tool name	Year of release	Key statistical method	Category
FunSpec	2002	Hypergeometric	Class I
Onto-express	2002	Fisher's exact; hypergeometic; binomial; chi-square	Class I
EASE	2003	Fisher's exact (modified as EASE score)	Class I
FatiGO/FatiWise/FatiGO+	2003	Fisher's exact	Class I
FuncAssociate	2003	Fisher's exact	Class I
GARBAN	2003	Hypergeometric	Class I
GeneMerge	2003	Hypergeometric	Class I
GoMiner	2003	Fisher's exact	Class I
MAPPFinder	2003	*Z*-score; hypergeometric	Class I
CLENCH	2004	Hypergeometric; chi-square; binomial	Class I
GO::TermFinder	2004	hypergeometric	Class I
GOAL	2004	Permutation	Class I
GOArray	2004	Hypergeometric; *Z*-score; permutation	Class I
GOStat	2004	Fisher's exact; chi-squre	Class I
GoSurfer	2004	Chi-square	Class I
OntologyTraverser	2004	Hypergeometric; Fisher's exact	Class I
THEA	2004	Hypergeometric	Class I
BiNGO	2005	Hypergeometric; binomial	Class I
FACT	2005	Adopt GeneMerge and GO::TermFinder statistical modules	Class I
gfinder	2005	Fisher's exact	Class I
Gobar	2005	Hypergeometric	Class I
GOCluster	2005	Hypergeometric	Class I
GOSSIP	2005	Fisher's exact	Class I
L2L	2005	Binomial; hypergeometric	Class I
WebGestalt	2005	Hypergeometric	Class I
BayGO	2006	Bayesian; Goodman and Kruskal's gamma factor	Class I
eGOn/GeneTools	2006	Fisher's exact	Class I
Gene Class Expression	2006	*Z*-statistics	Class I
GOALIE	2006	Hidden Kripke model	Class I
GOFFA	2006	Fisher's inverse chi-square	Class I
GOLEM	2006	Hyerpgeometric	Class I
JProGO	2006	Fisher's exact; Kolmogorov–Smirnov test; student's *t*-test; Wilcoxon's test; hypergeometric	Class I
PageMan	2006	Fisher's exact; chi-square; Wilcoxon	Class I
STEM	2006	Hypergeometric	Class I
WEGO	2006	Chi-square	Class I
EasyGO	2007	Hypergeometric; chi-square; binomial	Class I
g:Profiler	2007	Hypergeometric	Class I
ProbCD	2007	Yule's Q; Goodman-Kruskal's gamma; Cramer's T	Class I
GOEAST	2008	Hypergeometric	Class I
GOHyperGAll	2008	Hypergeometric	Class I
CatMap	2004	Permutations	Class II
Godist	2004	Kolmogorov–Smirnov test	Class II
GO-Mapper	2004	Gaussian distribution; EQ-score	Class II
iGA	2004	Permutations; hypergeometric; *t*-test; *Z*-score	Class II
GSEA	2005	Kolmogorov–Smirnov-like statistic	Class II
MEGO	2005	*Z*-score	Class II
PAGE	2005	*Z*-score	Class II
T-profiler	2005	*t*-Test	Class II
FuncCluster	2006	Fisher's exact	Class II
FatiScan	2007	Fisher's Exact	Class II
FINA	2007	Fisher's exact	Class II
GAzer	2007	*Z*-statistics; permutation	Class II
GeneTrail	2007	Hypergeometric; Kolmogorov–Smirnov	Class II
MetaGP	2007	*Z*-score	Class II
Ontologizer	2004	Fisher's exact	Class III
POSOC	2004	POSET (a discrete math: finite partially ordered set)	Class III
topGO	2006	Fisher's exact	Class III
GO-2D	2007	Hypergeometric; binomial	Class III
GENECODIS	2007	Hypergeometric; chi-square	Class III
GOSim	2007	Resnik's similarity	Class III
PalS	2008	Percent	Class III
ProfCom	2008	Greedy heuristics	Class III
GOTM	2004	Hypergeometric	Class I,II
ermineJ	2005	Permutations; Wilcoxon rank-sum test	Class I,II
DAVID	2003	Fisher's Exact (modified as EASE score)	Class I,III
GOToolBox	2004	Hypergeometric; Fisher's exact; Binomial	Class I,III
ADGO	2006	*Z*-statistic	Class II,III
FunNet	2008	Unclear	Unclear

Enrichment Analysis Tools

Table 2. Categorization of enrichment analysis tools

Tool category	Description	Indication and limitation	Sub-type of algorithms	Methods	Example tool
Class I: singular enrichment analysis (SEA)	Enrichment *P*-value is calculated on each term from the pre-selected interesting gene list. Then, enriched terms are listed in a simple linear text format. This strategy is the most traditional algorithm. It is still dominantly used by most of the enrichment analysis tools.	Capable of analyzing any gene list, which could be selected from any high-throughput biological studies/technologies (e.g. Microarray, ChIP-on-CHIP, ChIP-on-sequence, SNP array, EXON array, large scale sequence, etc.). However, the deeper inter-relationships among the terms may not be fully captured in linear format report.	Global reference background	Fisher's exact hypergeometric chi-square binomial	GoStat, GoMiner, GOTM, BinGO, GOtoolBox, GFinder, etc.
			Local reference background	Fisher's Exact hypergeometric chi-square binomial	DAVID, Onto-Express, GARBAN, FatiGO, etc.
			Neural network	Bayesian	BayGO
Class II: gene set enrichment analysis (GSEA)	Entire genes (without pre-selection) and associated experimental values are considered in the enrichment analysis. The unique features of this strategy are: (i) No need to pre-select interesting genes, as opposed to Classes I and II; (ii) Experimental values integrated into *P*-value calculation.	Suitable for pair-wide biological studies (e.g. disease versus control). Currently, may be difficult to be applied to the diverse data structures derived by a complex experimental design and some of the new technologies (e.g. SNP, EXON, Promoter arrays).	Based on ranked gene list	Kolmogorov–Smirnov-like	GSEA, CapMap, etc.
			Based on continuous gene values	*t*-Test permutation *Z*-score	FatiScan, ADGO, ermineJ, PAGE, iGA, GO-Mapper, GOdist, FINA, T-profiler, MetaGP, etc.
Class III: modular enrichment analysis (MEA)	This strategy inherits key spirit of SEA. However, the term–term/gene–gene relationships are considered into enrichment *P*-value calculation. The advantage of this strategy is that term–term/gene–gene relationship might contain unique biological meaning that is not held by a single term or gene. Such network/modular analysis is closer to the nature of biological data structure.	Capable of analyzing any gene lists, which could be selected from any high-throughput biological studies/technologies, like Class I. Emphasis on network relationships during analysis. 'Orphan' gene/term (with little relationships to other genes/terms), that sometimes could be very interesting, too, may be left out from the analysis.	Composite annotations	Measure enrichment on joint terms	ADGO, GeneCodis, ProfCom, etc.
			DAG Structure	Measure enrichment by considering parents-child relationships	topGO, Ontologizer, POSOC, etc.
			Global annotation relationship	Measure term–term global similarity with Kappa Statistics Czekanowski-Dice Pearson's correlation	DAVID, GoToolBox, etc.

Class I – SEA / ORA

- Evaluates the statistical significance on a fraction of genes in a particular pathway/gene set/annotation term
- 2x2 table method
- Fisher's exact test, chi-square test, binomial test, hypergeometric test
- Use p-value (and FDR) to sort the gene sets/pathways/annotation terms enriched in your gene list

Class II – GSEA / FCS

- Evaluates the statistical significance at particular pathway/gene set/annotation term instead of individual genes
- Covered in previous lecture GSEA
- Kolmogorov-Smirnov test, permutation
- Use p-value (and FDR) to sort the gene sets/pathways/annotation terms enriched from high-throughput “omics” data

Class III – MEA / PT

- Pathway topology (PT)-based or MEA methods have been developed to utilize the additional information.
- PT-based methods are essentially the same as FCS methods in that they perform the same three steps as FCS methods.
- The key difference between the two is the use of pathway topology to compute gene-level statistics.

Limitations

	Class I – ORA / SEA	**Class II – FCS / GSEA**	**Class III – MEA / PT**
Limitations	• Evaluate each gene independently • Different tests give different results that may alter the interpretation of the data • No interactions are considered	• Evaluate each pathway independently • No consideration of topology or interactions with the gene sets/pathways	• Very specific to cell type and cell-specific expression based on conditions • Topology dependency is hard to extract • Dynamic?

Recommendations / Rules

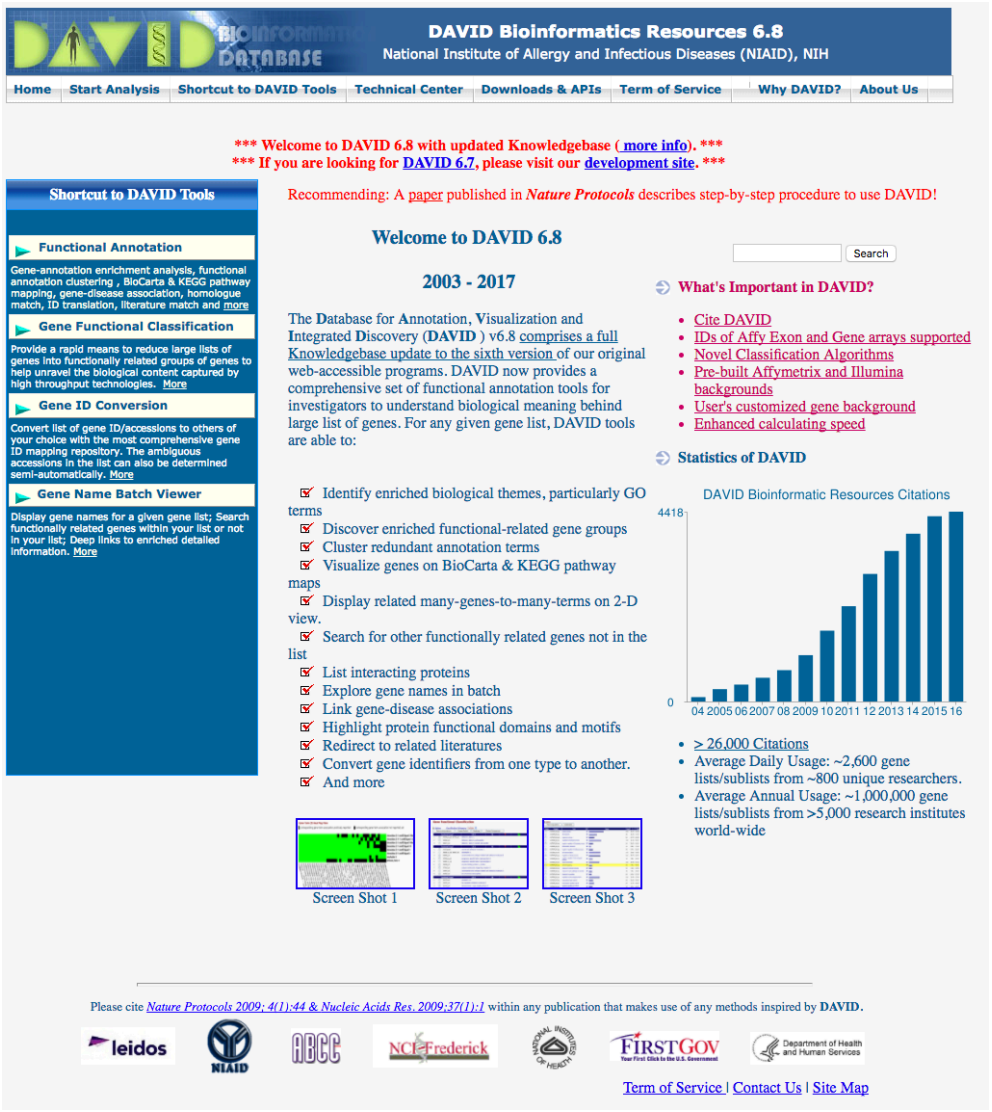
1. Realistically positioning the role of enrichment P-values in the current data-mining environment
2. Understanding the limitation of multiple testing correction on enrichment P-values
3. Cross-comparing enrichment analysis results derived from multiple gene lists

Recommendations / Rules

4. Setting up the ‘right’ gene reference background
5. Extending backend annotation databases
6. Efficiently mapping users’ input gene identifiers to the available annotation
7. Enhancing the exploratory capability and graphical presentation

https://david.ncifcrf.gov/

NIH DAVID



*** Welcome to DAVID 6.8 with updated Knowledgebase (more info). ***
*** If you are looking for DAVID 6.7, please visit our development site. ***

Upload | List | Background

Upload Gene List

Demolist 1 Demolist 2
Upload Help

Step 1: Enter Gene List

A: Paste a list

Clear

Or

B:Choose From a File

Choose File No file chosen

Multi-List File

Step 2: Select Identifier

AFFYMETRIX_3PRIME_IVT_ID

Step 3: List Type

Gene List
Background

Step 4: Submit List

Submit List

Functional Annotation Tool

Submit your gene list to start the tool!

Tell us how you like the tool
Read technical notes of the tool
Contact us for questions

Key Concepts:

Term/Gene Co-Occurrence Probability

Ranking functional categories based on co-occurrence with sets of genes in a gene list can rapidly aid in unraveling new biological processes associated with cellular functions and pathways. DAVID 6.8 allows investigators to sort gene categories from dozens of annotation systems. Sorting can be based either the number of genes within each category or by the EASE-score. More

Gene Similarity Search

Any given gene is associating with a set of annotation terms. If genes share similar set of those terms, they are most likely involved in similar biological mechanisms. The algorithm tries to group those related genes based on the agreement of sharing similar annotation terms by Kappa statistics. More

Term Similarity Search

Typically, a biological process/term is done by a corporation of a set of genes. If two or more biological processes are done by similar set of genes, the processes might be related in the biological network somehow. This search function is to identify the related biological processes/terms by quantitatively measuring the degree of the agreement how terms share the similar participating genes. More

Integrated Solutions

Functional Annotation Numerous Data Sources Co-occurrence Probability Use Homolog Annotation Dynamic Pathway Maps Disease Associations	Numerous public sources of protein and gene annotation have been parsed and integrated into DAVID 6.8. DAVID 6.8 contains information on over 1.5 million genes from more than 65,000 species. A list of protein or gene identifiers can be uploaded all at once to extract and summarize functional annotation associated with group of genes or with each individual gene. Data can be displayed in chart or table format or downloaded to the user?s hard drive.

*** Welcome to DAVID 6.8 with updated Knowledgebase (more info). ***
*** If you are looking for DAVID 6.7, please visit our development site. ***

Upload | List | Background

Gene List Manager

Select to limit annotations by one or more species Help

- Use All Species -
Homo sapiens(149)
Unknown(15)

Select Species

List Manager Help

demolist1

Select List to:

Use | Rename | Remove | Combine | Show Gene List

View Unmapped Ids

Annotation Summary Results

Help and Tool Manual

Current Gene List: demolist1 — **145 DAVID IDs**

Current Background: Homo sapiens — **Check Defaults** ☑ Clear All

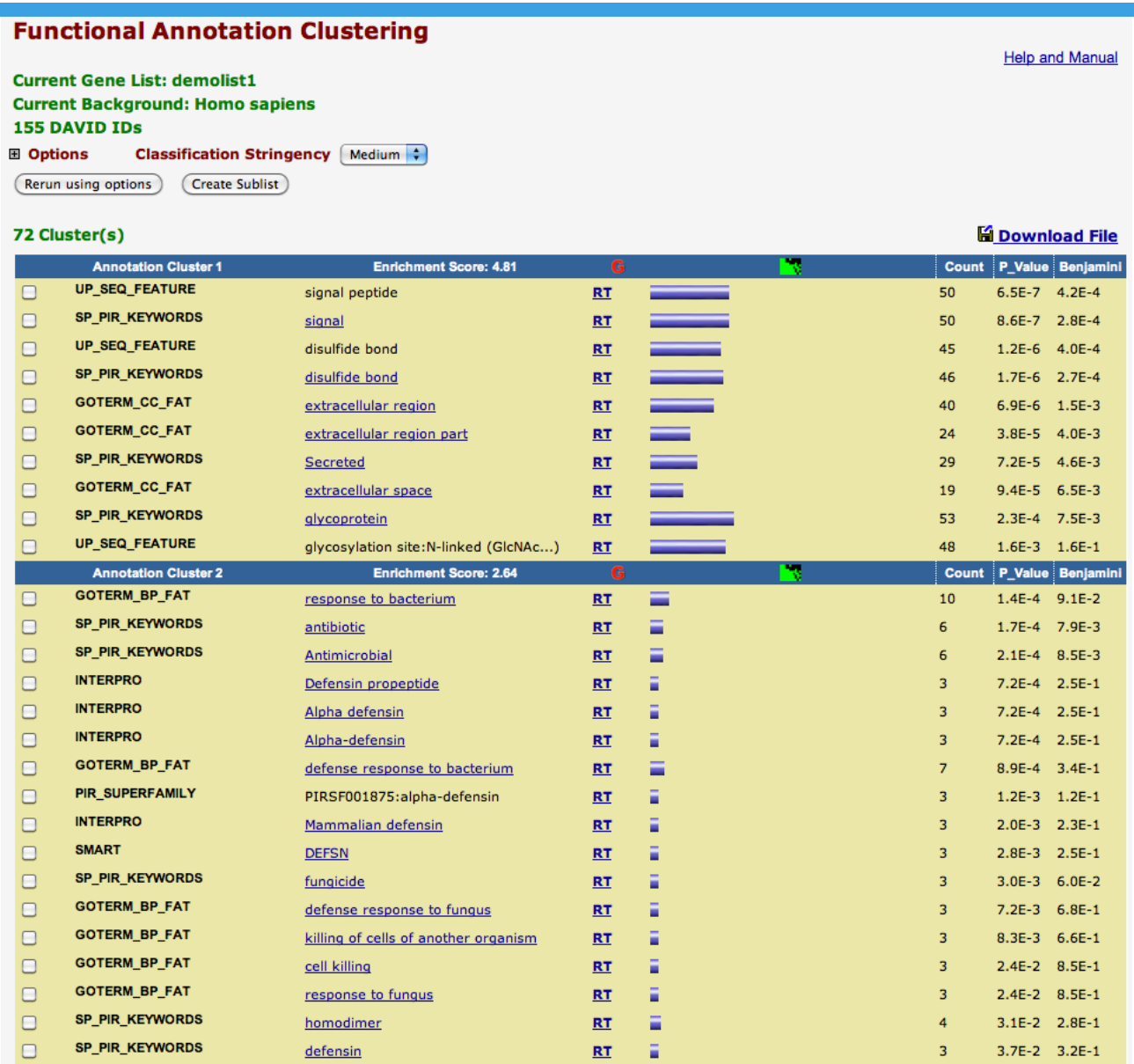
- ⊞ **Disease** (1 selected)
- ⊞ **Functional_Categories** (3 selected)
- ⊞ **Gene_Ontology** (3 selected)
- ⊞ **General_Annotations** (0 selected)
- ⊞ **Literature** (0 selected)
- ⊞ **Main_Accessions** (0 selected)
- ⊞ **Pathways** (3 selected)
- ⊞ **Protein_Domains** (3 selected)
- ⊞ **Protein_Interactions** (0 selected)
- ⊞ **Tissue_Expression** (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering
Functional Annotation Chart
Functional Annotation Table

Functional Annotation Clustering



Functional Annotation Clustering

Help and Manual

Current Gene List: demolist1
Current Background: Homo sapiens
155 DAVID IDs

⊞ Options Classification Stringency: Medium

Rerun using options Create Sublist

72 Cluster(s)

Download File

	Annotation Cluster 1	Enrichment Score: 4.81	G		Count	P_Value	Benjamini
☐	UP_SEQ_FEATURE	signal peptide	RT		50	6.5E-7	4.2E-4
☐	SP_PIR_KEYWORDS	signal	RT		50	8.6E-7	2.8E-4
☐	UP_SEQ_FEATURE	disulfide bond	RT		45	1.2E-6	4.0E-4
☐	SP_PIR_KEYWORDS	disulfide bond	RT		46	1.7E-6	2.7E-4
☐	GOTERM_CC_FAT	extracellular region	RT		40	6.9E-6	1.5E-3
☐	GOTERM_CC_FAT	extracellular region part	RT		24	3.8E-5	4.0E-3
☐	SP_PIR_KEYWORDS	Secreted	RT		29	7.2E-5	4.6E-3
☐	GOTERM_CC_FAT	extracellular space	RT		19	9.4E-5	6.5E-3
☐	SP_PIR_KEYWORDS	glycoprotein	RT		53	2.3E-4	7.5E-3
☐	UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	RT		48	1.6E-3	1.6E-1
	Annotation Cluster 2	**Enrichment Score: 2.64**	**G**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_FAT	response to bacterium	RT		10	1.4E-4	9.1E-2
☐	SP_PIR_KEYWORDS	antibiotic	RT		6	1.7E-4	7.9E-3
☐	SP_PIR_KEYWORDS	Antimicrobial	RT		6	2.1E-4	8.5E-3
☐	INTERPRO	Defensin propeptide	RT		3	7.2E-4	2.5E-1
☐	INTERPRO	Alpha defensin	RT		3	7.2E-4	2.5E-1
☐	INTERPRO	Alpha-defensin	RT		3	7.2E-4	2.5E-1
☐	GOTERM_BP_FAT	defense response to bacterium	RT		7	8.9E-4	3.4E-1
☐	PIR_SUPERFAMILY	PIRSF001875:alpha-defensin	RT		3	1.2E-3	1.2E-1
☐	INTERPRO	Mammalian defensin	RT		3	2.0E-3	2.3E-1
☐	SMART	DEFSN	RT		3	2.8E-3	2.5E-1
☐	SP_PIR_KEYWORDS	fungicide	RT		3	3.0E-3	6.0E-2
☐	GOTERM_BP_FAT	defense response to fungus	RT		3	7.2E-3	6.8E-1
☐	GOTERM_BP_FAT	killing of cells of another organism	RT		3	8.3E-3	6.6E-1
☐	GOTERM_BP_FAT	cell killing	RT		3	2.4E-2	8.5E-1
☐	GOTERM_BP_FAT	response to fungus	RT		3	2.4E-2	8.5E-1
☐	SP_PIR_KEYWORDS	homodimer	RT		4	3.1E-2	2.8E-1
☐	SP_PIR_KEYWORDS	defensin	RT		3	3.7E-2	3.2E-1

Functional Related Terms

Functional Related Terms

Options

Rerun using options

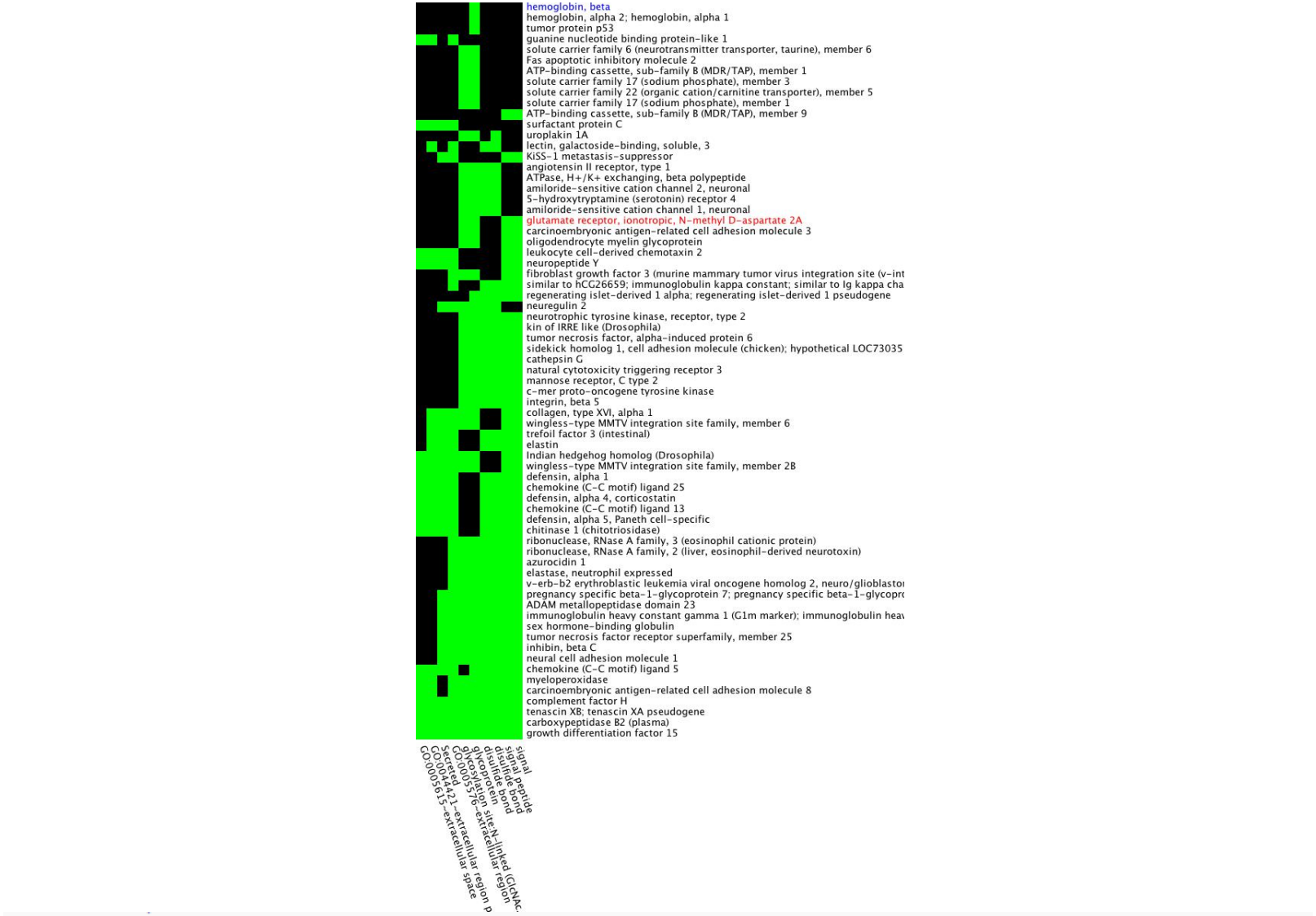
3896 term(s) were searched. 12 term(s) passed the filter.

Download File

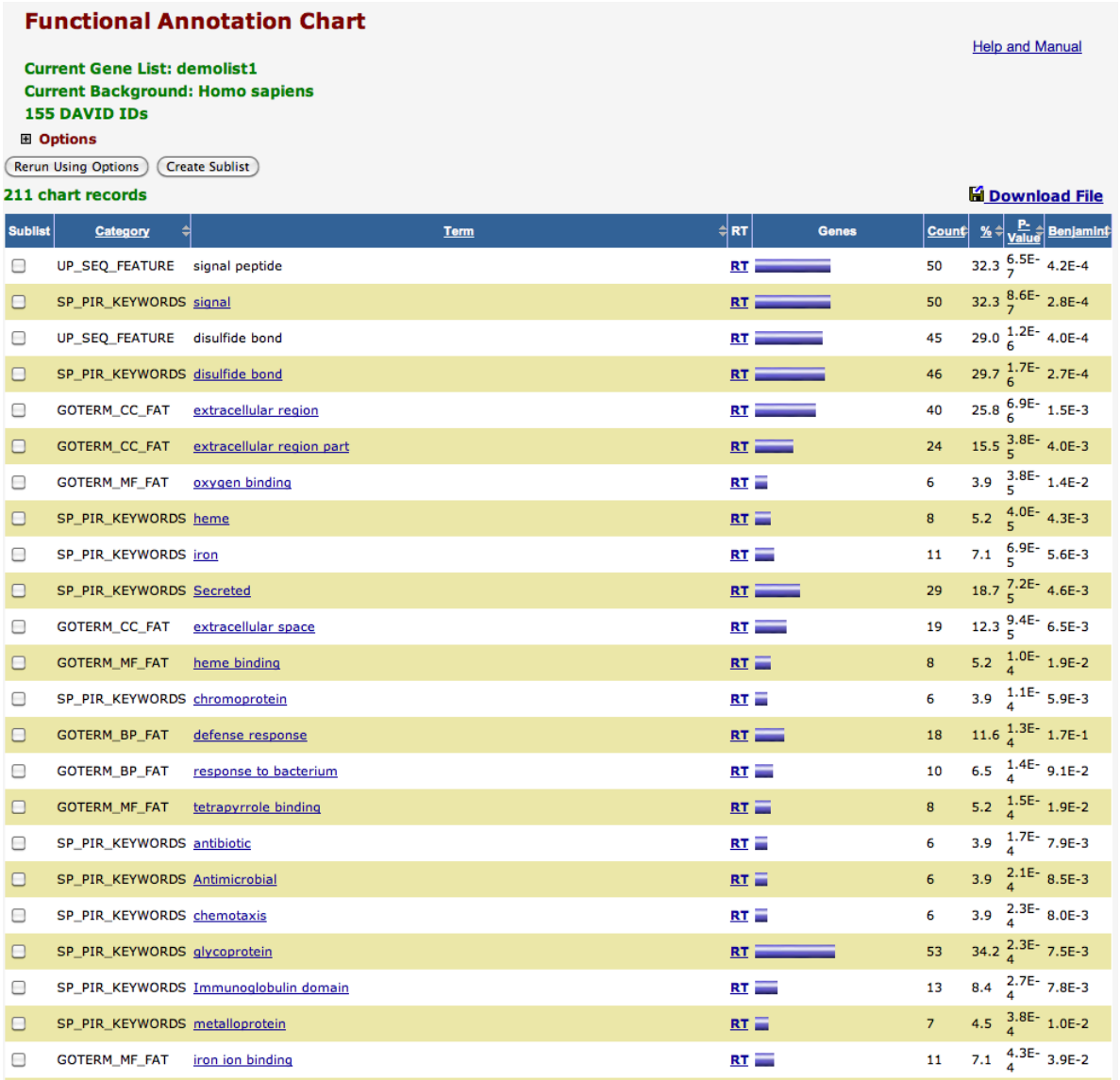
Similarity Score: Very High (0.75-1) High (0.5-0.75) Moderate (0.25-0.5) Low (<0.25)

#	Category	Term	Kappa
1	SP_PIR_KEYWORDS	signal	1.00
2	UP_SEQ_FEATURE	signal peptide	1.00
3	GOTERM_CC_FAT	extracellular region	0.72
4	UP_SEQ_FEATURE	disulfide bond	0.71
5	SP_PIR_KEYWORDS	disulfide bond	0.70
6	SP_PIR_KEYWORDS	Secreted	0.59
7	UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	0.58
8	SP_PIR_KEYWORDS	glycoprotein	0.58
9	GOTERM_CC_FAT	extracellular region part	0.45
10	GOTERM_CC_FAT	extracellular space	0.38
11	GOTERM_BP_FAT	defense response	0.33
12	GOTERM_BP_FAT	cell surface receptor linked signal transduction	0.30

2D View of Functional Annotation Clusters



Functional Annotation Chart



Functional Annotation Chart

Help and Manual

Current Gene List: demolist1
Current Background: Homo sapiens
155 DAVID IDs

⊞ Options

Rerun Using Options | Create Sublist

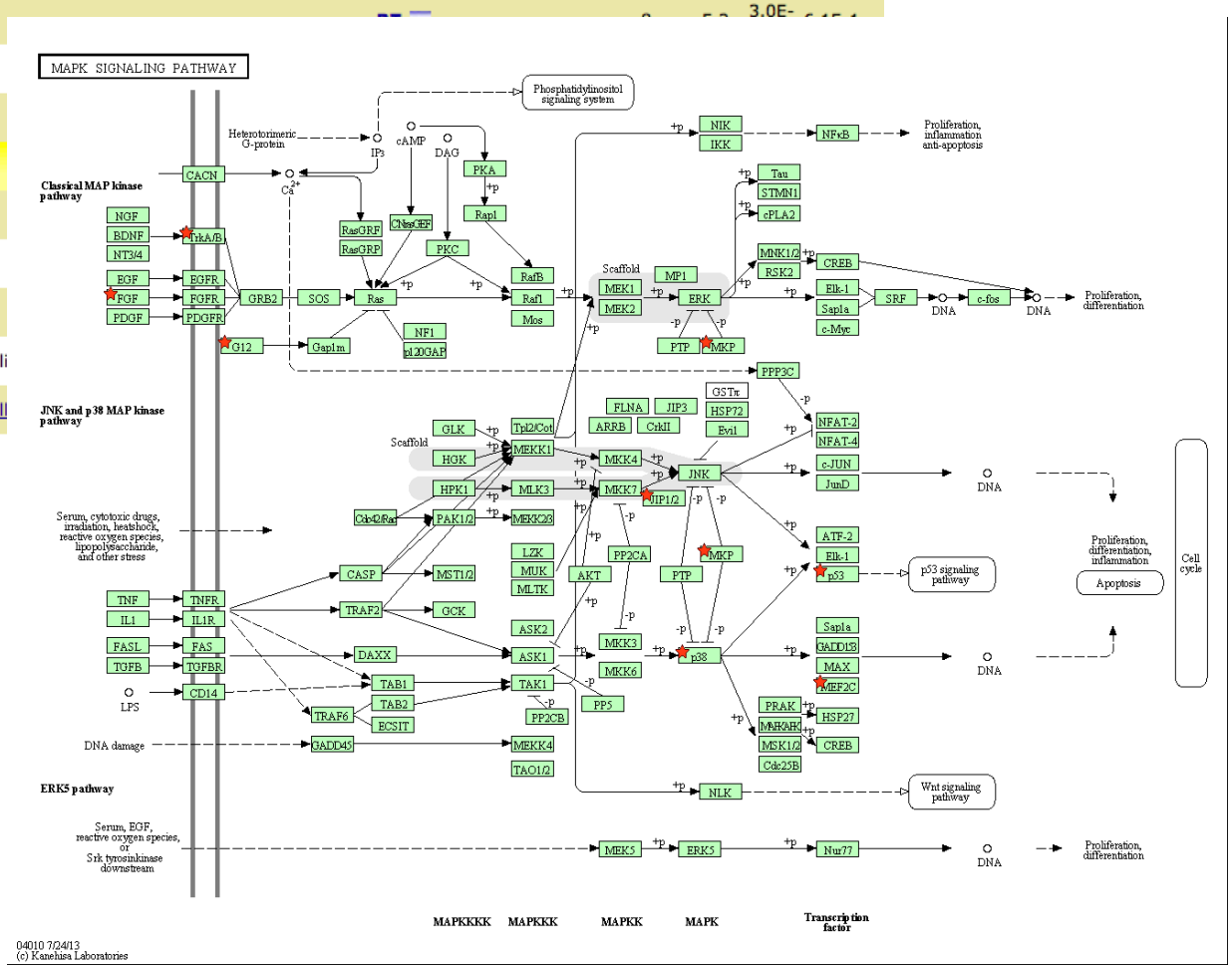
211 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
☐	UP_SEQ_FEATURE	signal peptide	RT		50	32.3	6.5E-7	4.2E-4
☐	SP_PIR_KEYWORDS	signal	RT		50	32.3	8.6E-7	2.8E-4
☐	UP_SEQ_FEATURE	disulfide bond	RT		45	29.0	1.2E-6	4.0E-4
☐	SP_PIR_KEYWORDS	disulfide bond	RT		46	29.7	1.7E-6	2.7E-4
☐	GOTERM_CC_FAT	extracellular region	RT		40	25.8	6.9E-6	1.5E-3
☐	GOTERM_CC_FAT	extracellular region part	RT		24	15.5	3.8E-5	4.0E-3
☐	GOTERM_MF_FAT	oxygen binding	RT		6	3.9	3.8E-5	1.4E-2
☐	SP_PIR_KEYWORDS	heme	RT		8	5.2	4.0E-5	4.3E-3
☐	SP_PIR_KEYWORDS	iron	RT		11	7.1	6.9E-5	5.6E-3
☐	SP_PIR_KEYWORDS	Secreted	RT		29	18.7	7.2E-5	4.6E-3
☐	GOTERM_CC_FAT	extracellular space	RT		19	12.3	9.4E-5	6.5E-3
☐	GOTERM_MF_FAT	heme binding	RT		8	5.2	1.0E-4	1.9E-2
☐	SP_PIR_KEYWORDS	chromoprotein	RT		6	3.9	1.1E-4	5.9E-3
☐	GOTERM_BP_FAT	defense response	RT		18	11.6	1.3E-4	1.7E-1
☐	GOTERM_BP_FAT	response to bacterium	RT		10	6.5	1.4E-4	9.1E-2
☐	GOTERM_MF_FAT	tetrapyrrole binding	RT		8	5.2	1.5E-4	1.9E-2
☐	SP_PIR_KEYWORDS	antibiotic	RT		6	3.9	1.7E-4	7.9E-3
☐	SP_PIR_KEYWORDS	Antimicrobial	RT		6	3.9	2.1E-4	8.5E-3
☐	SP_PIR_KEYWORDS	chemotaxis	RT		6	3.9	2.3E-4	8.0E-3
☐	SP_PIR_KEYWORDS	glycoprotein	RT		53	34.2	2.3E-4	7.5E-3
☐	SP_PIR_KEYWORDS	Immunoglobulin domain	RT		13	8.4	2.7E-4	7.8E-3
☐	SP_PIR_KEYWORDS	metalloprotein	RT		7	4.5	3.8E-4	1.0E-2
☐	GOTERM_MF_FAT	iron ion binding	RT		11	7.1	4.3E-4	3.9E-2

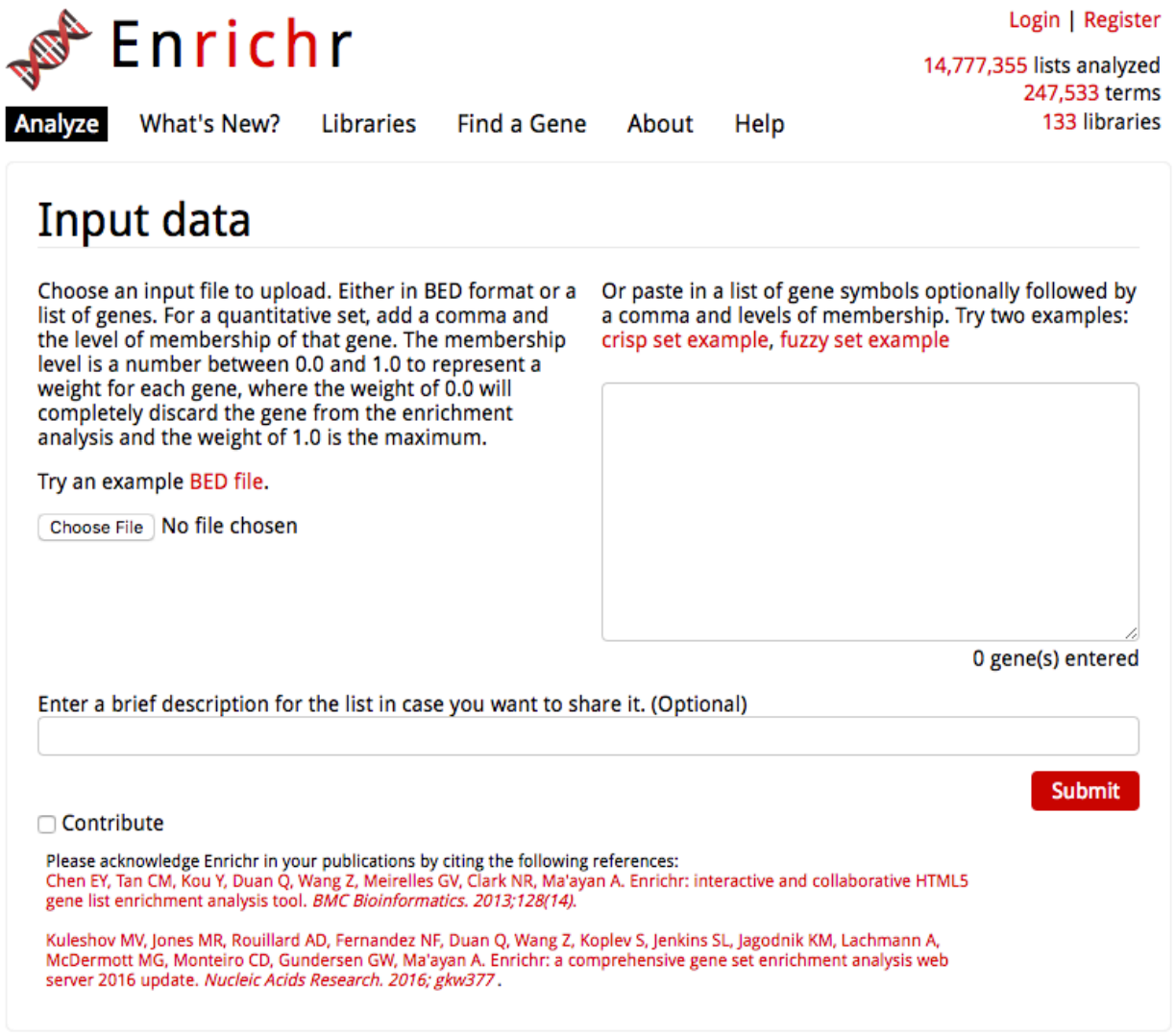
Linking Enriched Genes to Pathways (KEGG only)

☐	GOTERM_BP_FAT	oxygen and reactive oxygen species metabolic process	RT	4	2.6	2.9E-2	8.3E-1
☐	UP_SEQ_FEATURE	short sequence motif:Cell attachment site	RT	4	2.6	2.9E-2	7.0E-1
☐	GOTERM_MF_FAT	ATPase activity				3.0E-	
☐	GOTERM_BP_FAT	cellular amino acid catabolic process					
☐	SP_PIR_KEYWORDS	homodimer					
☐	KEGG_PATHWAY	MAPK signaling pathway					
☐	GOTERM_MF_FAT	ATPase activity, coupled					
☐	GOTERM_MF_FAT	ErbB-3 class receptor binding					
☐	SP_PIR_KEYWORDS	defensin					
☐	UP_SEQ_FEATURE	metal ion-binding site:Iron (heme axial li					
☐	GOTERM_BP_FAT	positive regulation of striated muscle cell					



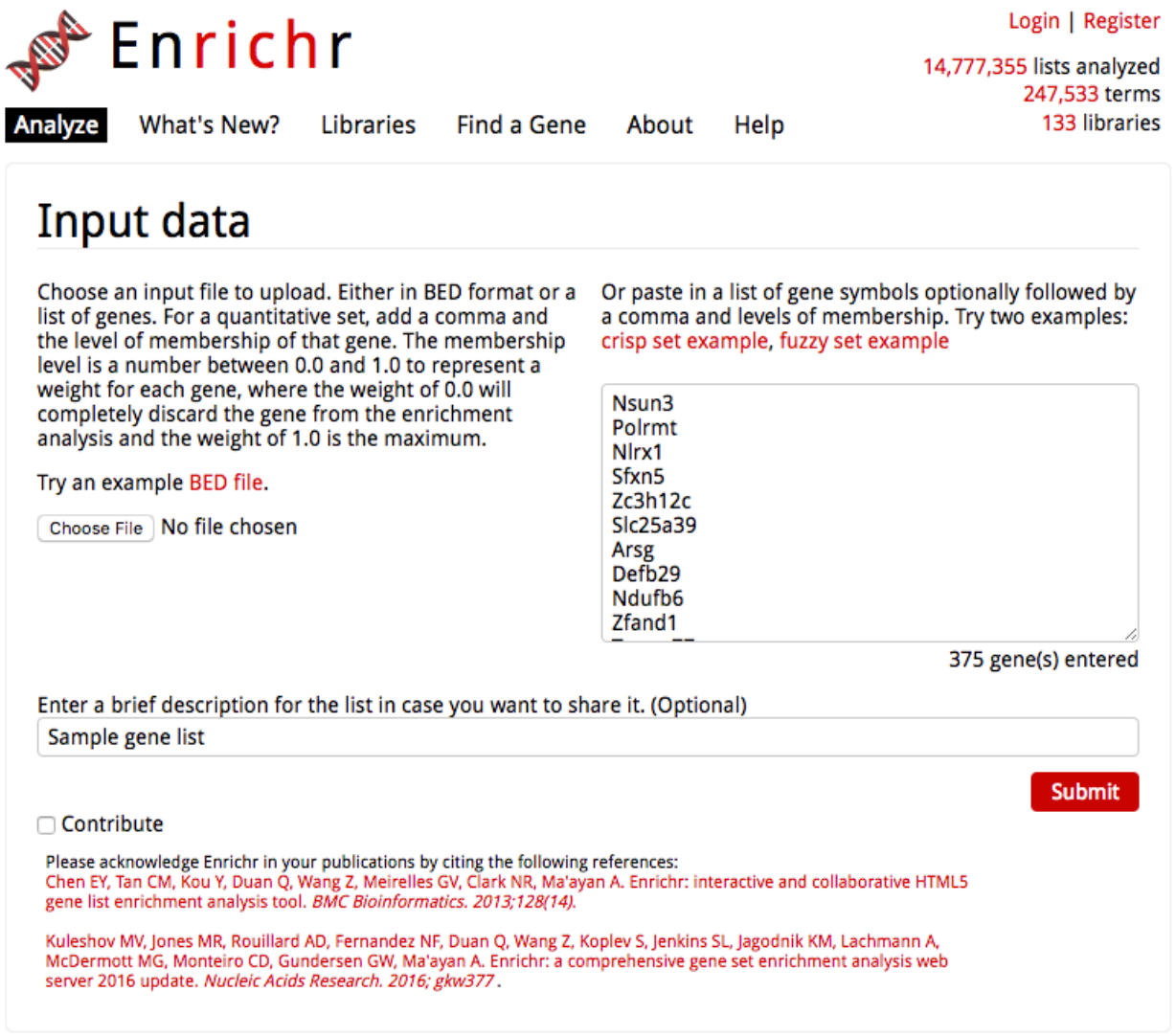
Enrichr

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



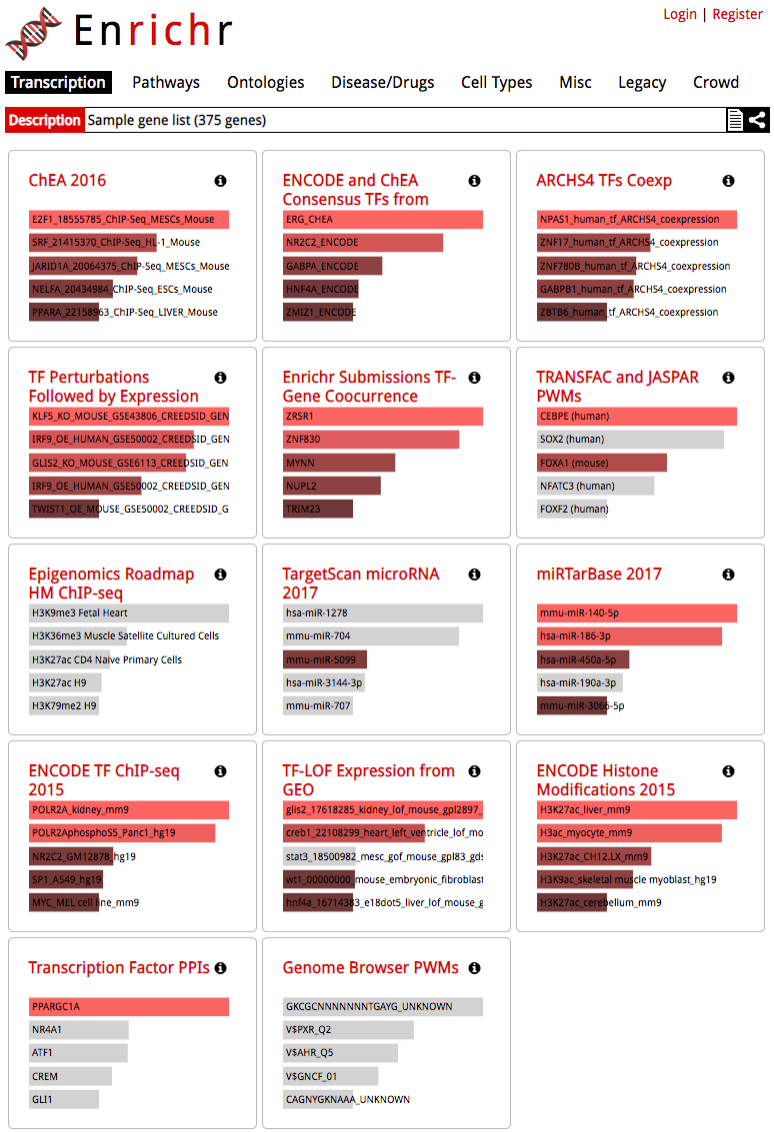
Enrichr

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



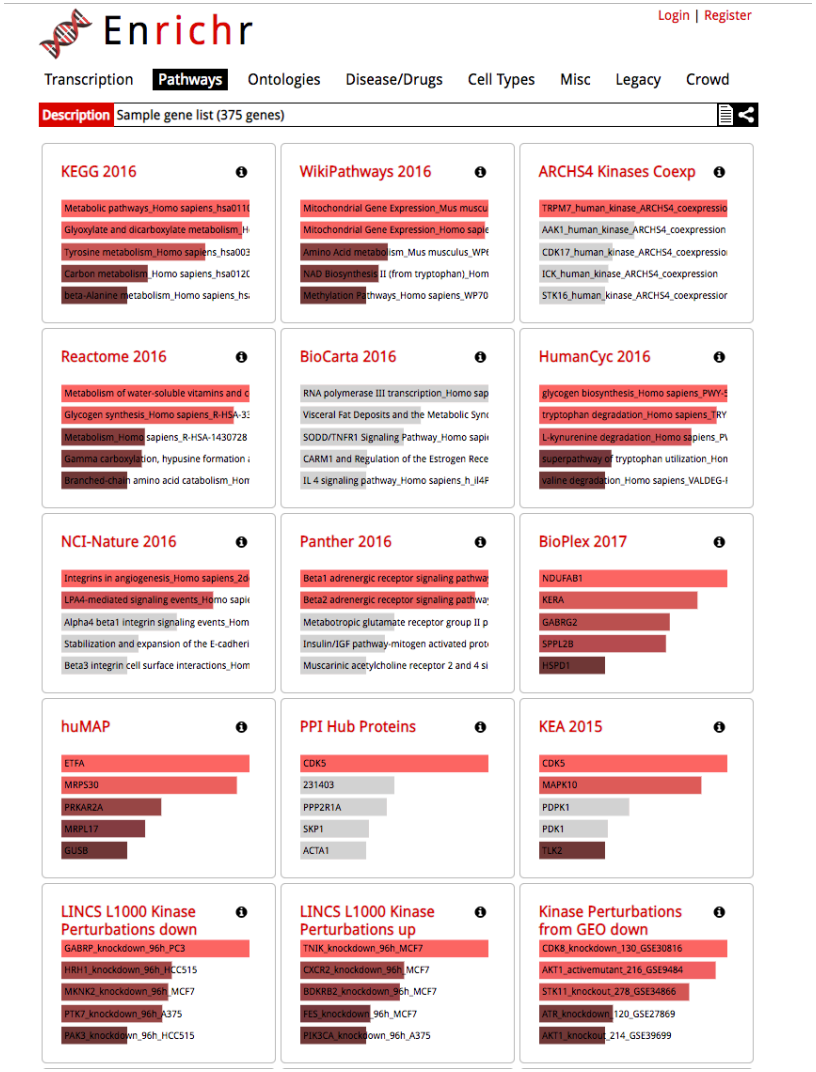
Enrichr

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



Enrichr

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



Enrichr

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

WEBGESTAL

WEB-based GEne SeT AnaLysis Toolkit

WebGestalt *Translating gene lists into biological insights...*

http://www.webgestalt.org/option.php

ORA Sample Run | GSEA Sample Run | NTA Sample Run | External Examples | Manual | Citation | User Forum

GOView | WebGestaltR | WebGestalt 2013

» Basic Parameters

Select Organism of Interest -- Organisms --

Select Method of Interest -- Methods --

Select Functional Database -- Functional Database Class --

-- Functional Database Name --

Gene List

Select Gene ID Type -- Gene ID Type --

Choose File No file chosen Reset

Upload Gene List (max size: 5 MB) OR

Please enter gene ids... Clear

Reference Gene List

-- Reference Gene Set --

Select Reference Set for Enrichment Analysis Reset

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type -- Reference Gene ID Type --

Choose File No file chosen Reset

» Advanced parameters

Submit

» Introduction

WebGestalt (WEB-based Gene SeT AnaLysis Toolket) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

» Data Source

» News

Browser support: PC: Google Chrome 56.0 or later, Mac: Google Chrome 56.0, Safari 10.0 or later. We strongly recommend upgrading to the latest version of the supported broswers.
For Safari users, please enable Flash for network visualization. Detailed information on how to enable Flash can be found here.

WebGestalt is currently developed and maintained by Jing Wang, Suhas Vasaikar, Zhiao Shi and Bing Zhang at the Zhang Lab. Other people who have made significant contribution to the project include Dexter Duncan, Stefan Kirov and Jay Snoddy.

Funding credits: NIH/NCI (U24 CA210954); Leidos (15X038); CPRIT (RR160027); NIH/NIAAA (U01 AA016662, U01 AA013512); NIH/NIDA (P01 DA015027); NIH/NIMH (P50 MH078028, P50 MH096972); NIH/NCI (U24 CA159988); NIH/NIGMS (R01 GM088822).

» Basic Parameters

Select Organism of Interest hsapiens

Select Method of Interest Overrepresentation Enrichment Analysis (ORA)

Select Functional Database geneontology

Biological_Process

Gene List

Select Gene ID Type genesymbol

Choose File No file chosen Reset

Upload Gene List (max size: 5 MB) **OR**

ABCA1
ABCC9
ABCE1

Clear

Reference Gene List

genome_protein-coding

Select Reference Set for Enrichment Analysis Reset

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type -- Reference Gene ID Type --

Choose File No file chosen Reset

WEBGESTAL

Summary | User ID Mapping Table | GOSlim Summary | Enrichment Results

Summary (Result Download)

Enrich method: ORA
Organism:hsapiens
Enrichment Categories: geneontology_Biological_Process
Interesting gene list: textAreaUpload_1506444918.txt. **ID type:** genesymbol
The interesting gene list contains **487** user IDs in which **478** user IDs are unambiguously mapped to the unique Entrez Gene IDs and **9** user IDs are mapped to multiple Entrez Gene IDs or could not be mapped to any Entrez Gene ID. The GO Slim summary are based upon the **478** unique Entrez Gene IDs.
Among the **478** unique Entrez Gene IDs, **464** IDs are annotated to the selected functional categories and also in the reference gene list, which are used for the enrichment analysis.
Reference gene list: all mapped Entrez Gene IDs from the selected platform genome_protein-coding
The reference gene list contains **20691** IDs and **16258** IDs are annotated to the selected functional categories that are used as the reference for the enrichment analysis.
Parameters for the enrichment analysis:

- **Minimum number of Entrez Gene IDs in the category:**5
- **Maximum number of Entrez Gene IDs in the category:**2000
- **FDR Method:**BH
- **Significance Level:** Top10

Based on the above parameters, **10** categories are identified as enriched categories and all are shown in this report.

WEBGESTAL

Summary | **User ID Mapping Table** | GOSlim Summary | Enrichment Results

Mapped User IDs

userid	Gene Symbol	Gene Name	Entrez Gene
CD24	CD24	CD24 molecule	100133941
ABCC9	ABCC9	ATP binding cassette subfamily C member 9	10060
PTPRU	PTPRU	protein tyrosine phosphatase, receptor type U	10076
PQBP1	PQBP1	polyglutamine binding protein 1	10084
AKAP9	AKAP9	A-kinase anchoring protein 9	10142
CDH17	CDH17	cadherin 17	1015
DHRS2	DHRS2	dehydrogenase/reductase 2	10202
CDK8	CDK8	cyclin dependent kinase 8	1024
EFS	EFS	embryonal Fyn-associated substrate	10278
BCAS2	BCAS2	BCAS2, pre-mRNA processing factor	10286
SF3A1	SF3A1	splicing factor 3a subunit 1	10291
CNPY2	CNPY2	canopy FGF signaling regulator 2	10330
TFG	TFG	TRK-fused gene	10342
CITED2	CITED2	Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2	10370
TIMM17A	TIMM17A	translocase of inner mitochondrial membrane 17A	10440
MCRS1	MCRS1	microspherule protein 1	10445
ZNHIT1	ZNHIT1	zinc finger HIT-type containing 1	10467

User IDs mapped to multiple Entrtez IDs or not mapped

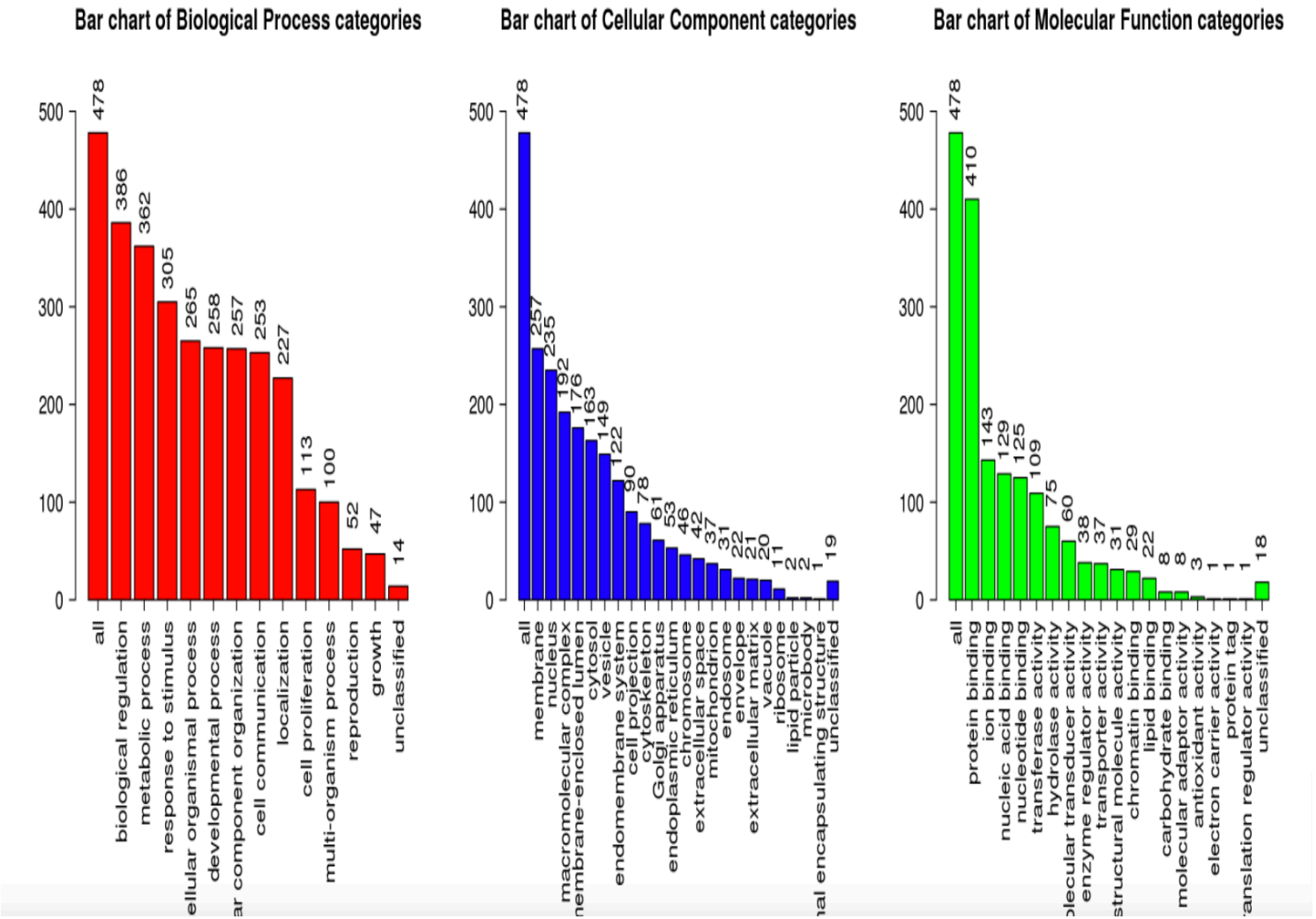
userid
ADRBK1
C12orf11
CXCR7
DFNB31
KIAA0182
MLL3
NHP2L1
PAK7
SNHG3-RCC1

WEBGESTAL

Summary | User ID Mapping Table | GOSlim Summary | Enrichment Results

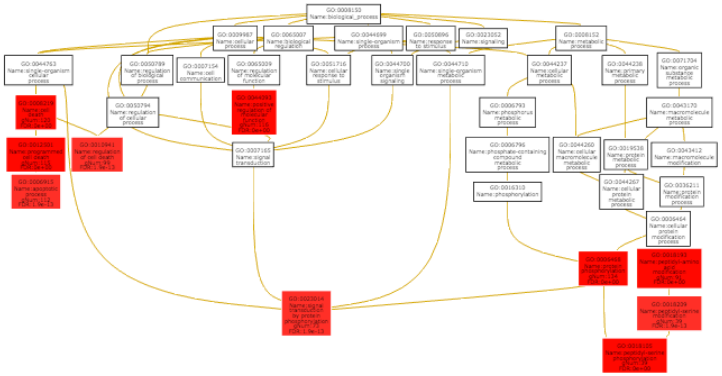
GOSlim summary for the user list genes

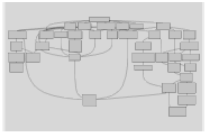
Each Biological Process, Cellular Component and Molecular Function category is represented by a red, blue and green bar, repectively.
The height of the bar represents the number of user list genes observed in the category.



WEBGESTAL

Summary | User ID Mapping Table | GOSlim Summary | **Enrichment Results**





Detailed information of the enriched categories

The statistics for the enriched categories and the genes in the user gene list and also in the category are listed in the table.

ID Search

Download Table

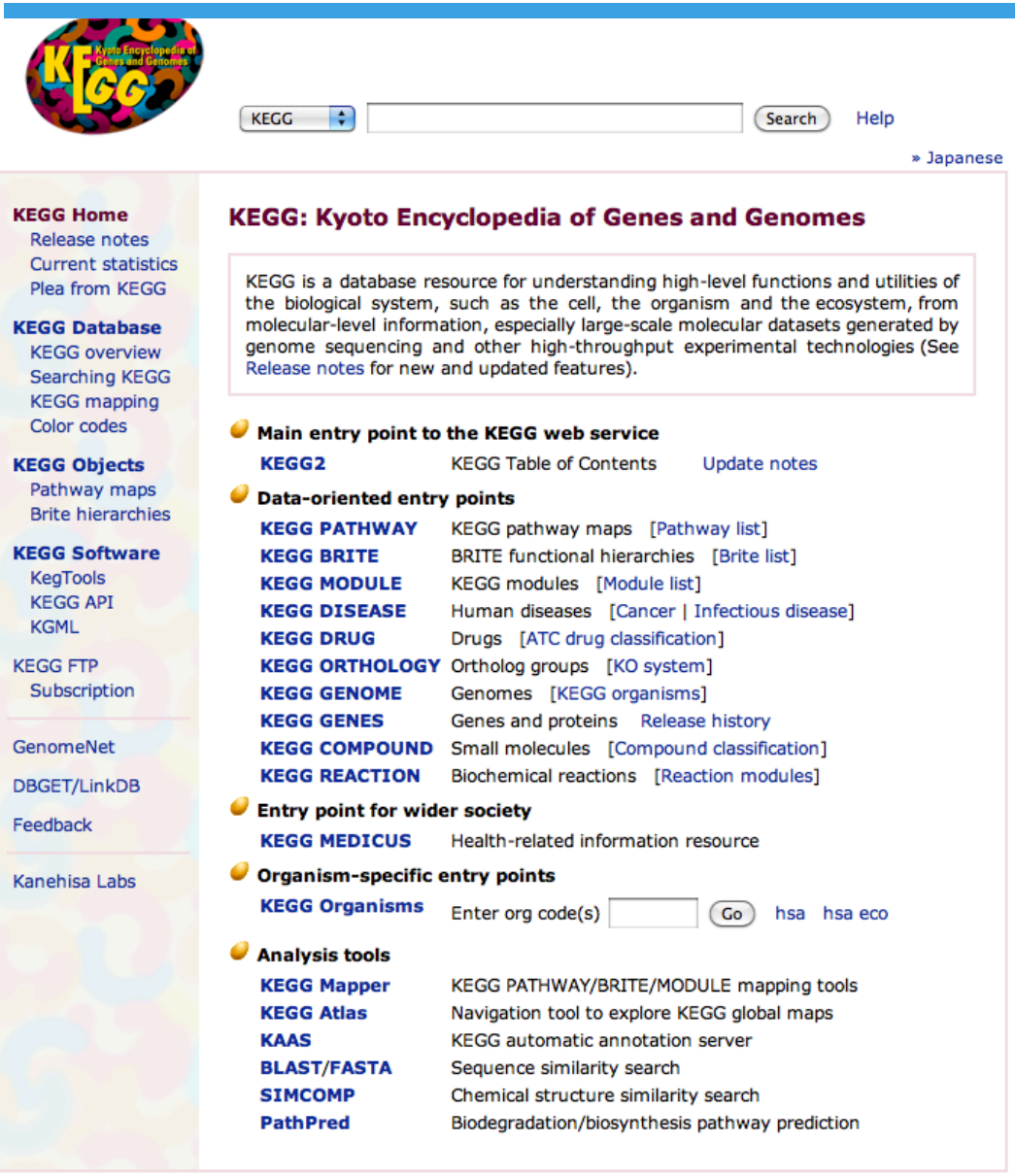
ID:GO:0006468 Name:protein phosphorylation

C=1824; O=134; E=51.72; R=2.59; PValue=0e+00; FDR=0e+00

userid	Gene Symbol	Gene Name	Entrez Gene
CD24	CD24	CD24 molecule	100133941
AKAP9	AKAP9	A-kinase anchoring protein 9	10142
CDK8	CDK8	cyclin dependent kinase 8	1024
CAMKK2	CAMKK2	calcium/calmodulin dependent protein kinase kinase 2	10645
COPS8	COPS8	COP9 signalosome subunit 8	10920
PWP1	PWP1	PWP1 homolog, endonuclein	11137
CHEK2	CHEK2	checkpoint kinase 2	11200
IRAK3	IRAK3	interleukin 1 receptor associated kinase 3	11213
ADCY8	ADCY8	adenylate cyclase 8	114
CHUK	CHUK	conserved helix-loop-helix ubiquitous kinase	1147
CSNK1A1L	CSNK1A1L	casein kinase 1 alpha 1 like	122011
CNTN1	CNTN1	contactin 1	1272
ADORA1	ADORA1	adenosine A1 receptor	134

KEGG

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



KEGG

KEGG2 PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET

Search KEGG for Go Clear

Category	Entry Point	Search & Compute	DBGET Search
Systems information	**KEGG PATHWAY** **KEGG BRITE** **KEGG MODULE** **KEGG Mapper** **KEGG Atlas**	Search Pathway Search Brite Reconstruct Module Map Taxonomy	PATHWAY BRITE MODULE
Genomic information	**KEGG ORTHOLOGY** **KEGG Annotation**	BlastKOALA *New!* KO system	ORTHOLOGY
	KEGG GENOME **KEGG GENES** **KEGG Organisms** [Species | Genus]	SSDB search OC viewer† BLAST† / FASTA† KAAS†	GENOME GENES DGENES MGENOME† MGENES†
Chemical information	**KEGG LIGAND** **KEGG COMPOUND** **KEGG GLYCAN** **KEGG REACTION** **Reaction Modules**	SIMCOMP† / SUBCOMP† KCaM† PathSearch† PathComp† PathPred† E-zyme†	COMPOUND GLYCAN REACTION RPAIR RCLASS ENZYME
Health information	**KEGG DISEASE** **KEGG DRUG** **KEGG ENVIRON** **KEGG MEDICUS**	MEDICUS search Drug interaction checker Human diseases Infectious diseases ATC drug classification	DISEASE DRUG DGROUP ENVIRON

KEGG is developed by Kanehisa Laboratories. See Kanehisa et al. (2014) for updates of KEGG.
† Developed and maintained by Kyoto University Bioinformatics Center as part of its GenomeNet service.

KEGG

KEGG2 PATHWAY BRITE MODULE DISEASE DRUG KO GENOME GENES LIGAND DBGET

Select prefix: map | Organism

Enter keywords: | Go | Help

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps and update history) representing our knowledge on the molecular interaction and reaction 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**

and also on the structure relationships (KEGG drug structure maps) in:

7. **Drug Development**

Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretaion of higher-level systemic functions.

- Search Pathway - basic pathway mapping tool
- Search&Color Pathway - advanced pathway mapping tool
- Color Pathway - selected pathway map coloring tool

About KEGG Atlas

KEGG Pathway Maps

map | ko | rn | hsa | eco | edit

Metabolism

01100 Metabolic pathways

01110 Biosynthesis of secondary metabolites

01120 Microbial metabolism in diverse envir

01200 Carbon metabolism

01210 2-Oxocarboxylic acid metabolism

01230 Biosynthesis of amino acids

01220 Degradation of aromatic compounds

Layer

KEGG Module

check / uncheck

Energy metabolism

Carbon fixation

M00165 Reductive pentose phosphate c

M00166 Reductive pentose phosphate c

M00167 Reductive pentose phosphate c

M00168 CAM (Crassulacean acid metab

M00169 CAM (Crassulacean acid metab

M00172 C4-dicarboxylic acid cycle, NAD

M00171 C4-dicarboxylic acid cycle, NAD

M00170 C4-dicarboxylic acid cycle, phos

M00173 Reductive citrate cycle (Arnon-E

M00579 Phosphate acetyltransferase-ac

Nitrogen metabolism

M00175 Nitrogen fixation, nitrogen => ar

Methane metabolism

M00567 Methanogenesis, CO2 => meth

M00174 Methane oxidation, methanotrop

Sulfur metabolism

M00176 Assimilatory sulfate reduction, s

Carbohydrate and lipid metabolism

Central carbohydrate metabolism

M00001 Glycolysis (Embden-Meyerhof p

M00002 Glycolysis, core module involvi

M00003 Gluconeogenesis, oxaloacetate

M00307 Pyruvate oxidation, pyruvate =>

M00009 Citrate cycle (TCA cycle, Krebs

M00010 Citrate cycle, first carbon oxidat

M00011 Citrate cycle, second carbon ox

M00004 Pentose phosphate pathway (P

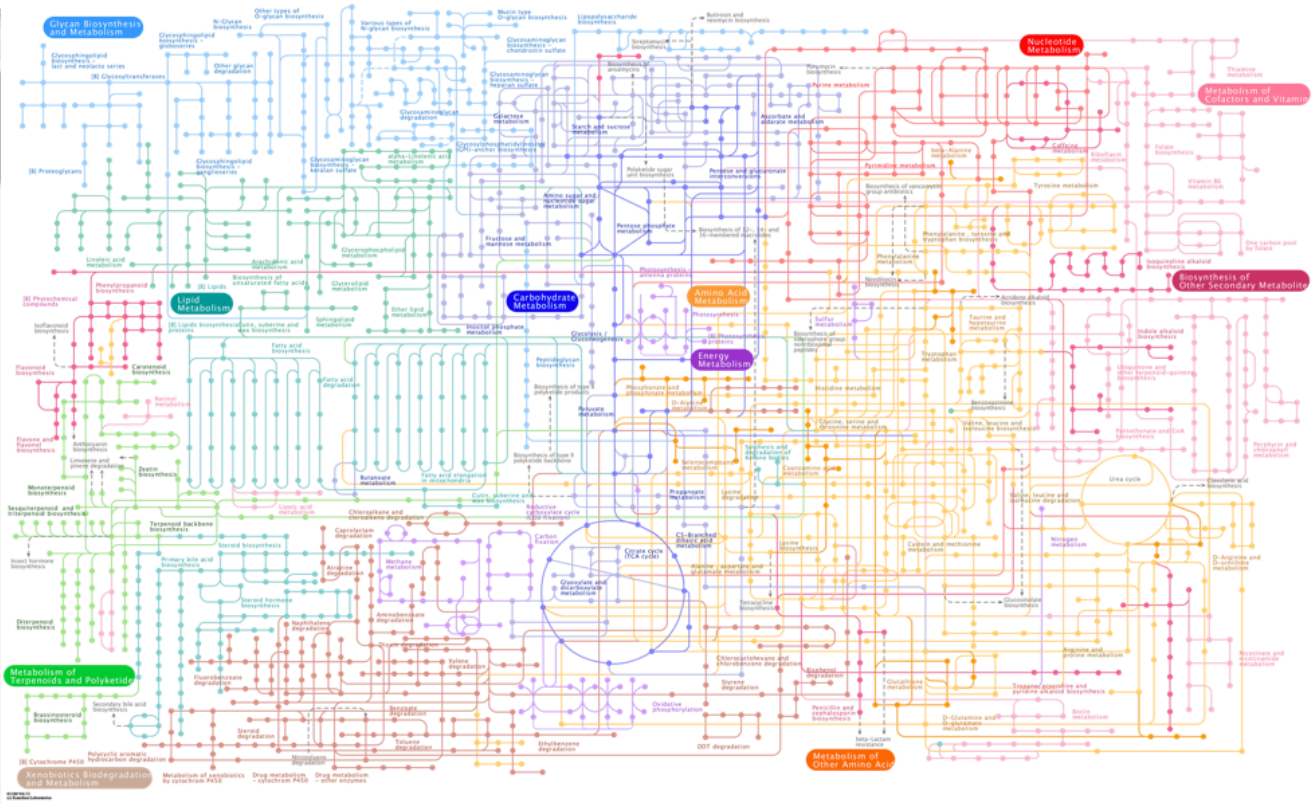
M00006 Pentose phosphate pathway, ox

M00007 Pentose phosphate pathway, nc

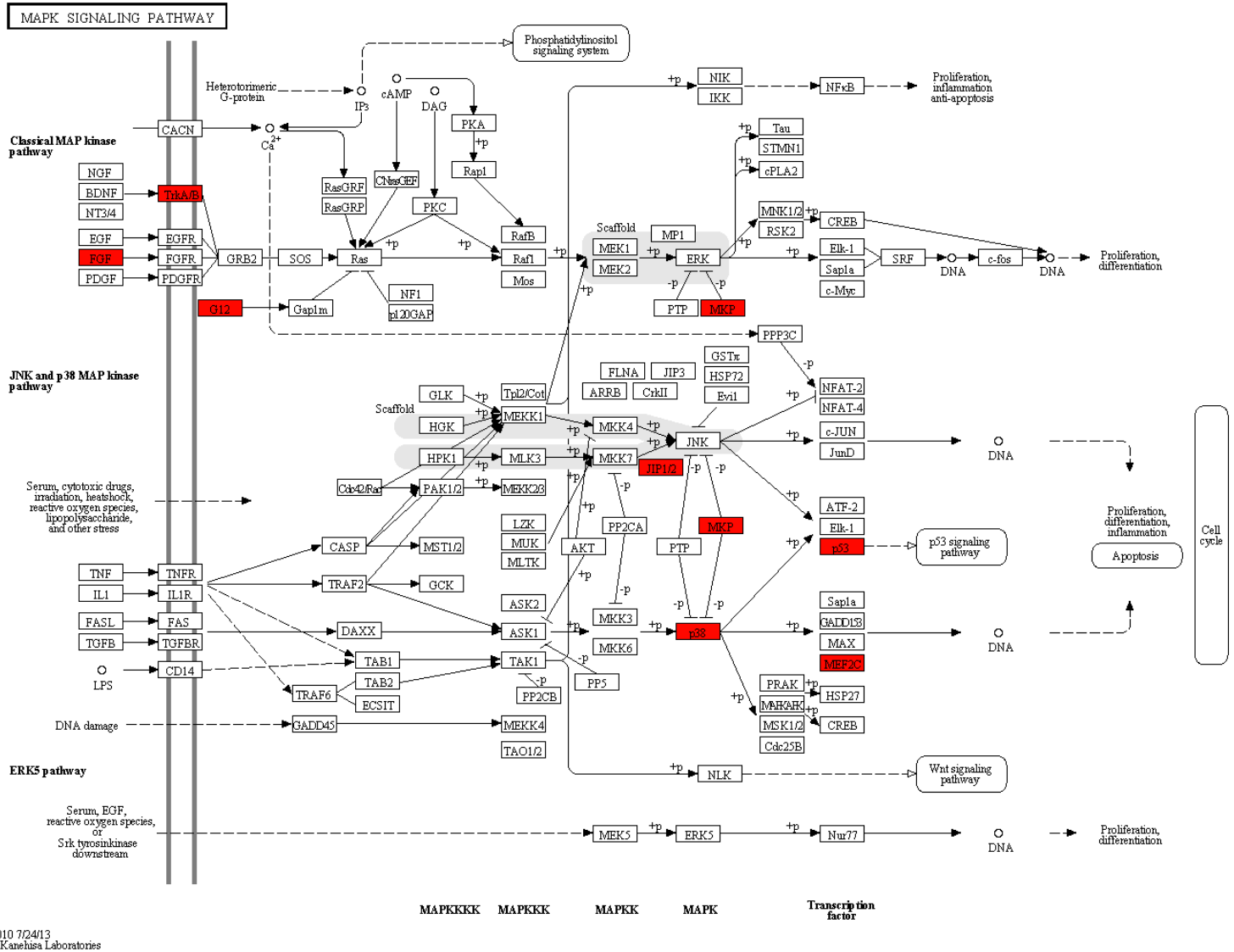
Search Result

History

KEGG



KEGG

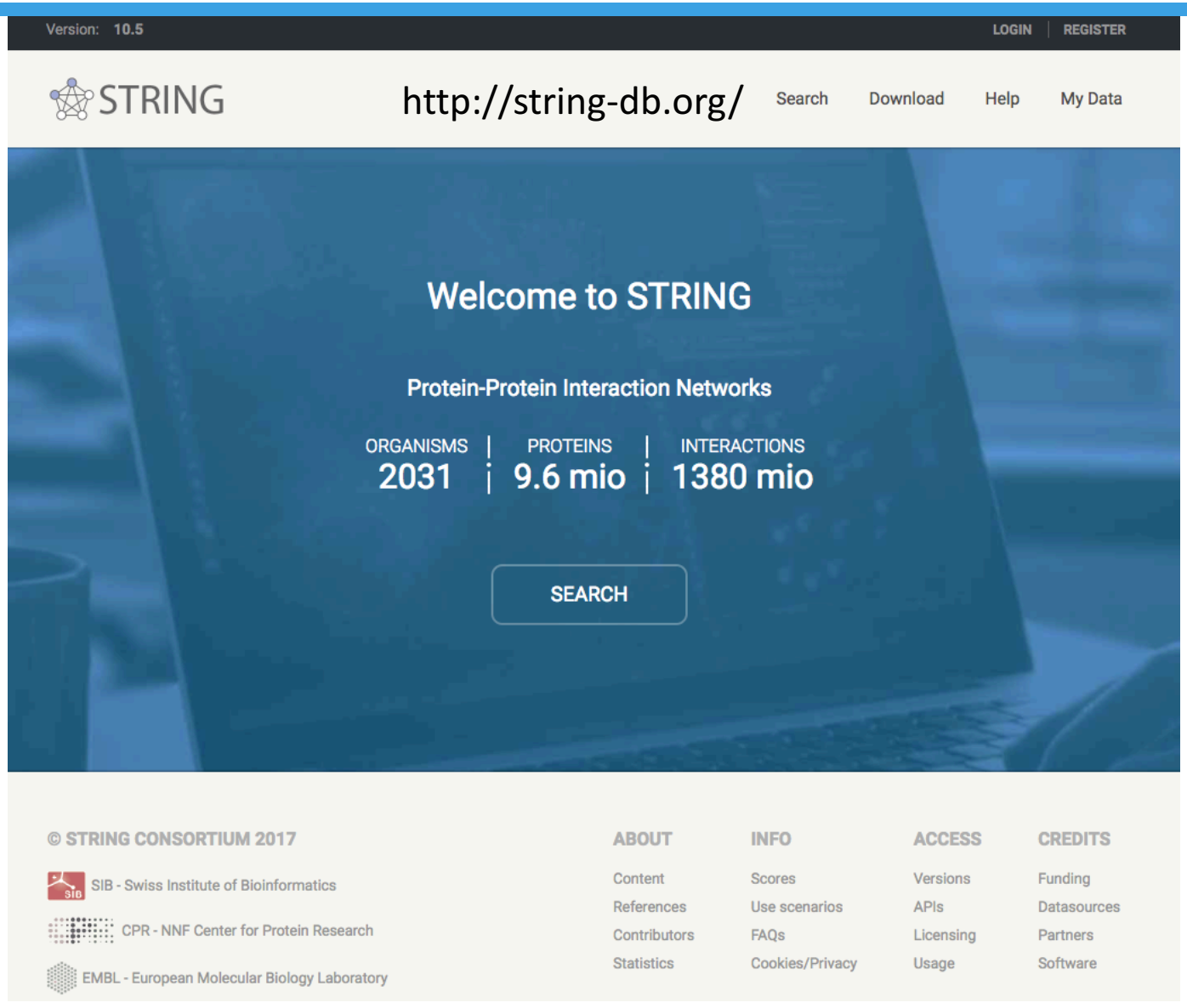

MAPK SIGNALING PATHWAY
Classical MAP kinase pathway
JNK and p38 MAP kinase pathway
ERK5 pathway
Phosphatidylinositol signaling system
Heterotorimeric G-protein
cAMP
DAG
IP3
Ca2+
CACN
PKA
Rap1
RasGRF
CrkIIGEF
PKC
NGF
BDNF
NT3/4
TrkA/B
EGF
FGF
PDGF
EGFR
FGFR
PDGFR
GRB2
SOS
Ras
RafB
Raf1
Mos
NF1
pl20GAP
G12
Gaplm
Scaffold
MEK1
MEK2
MP1
ERK
PTP
MKP
NIK
IKK
NFκB
Proliferation, inflammation anti-apoptosis
Tau
STMN1
cPLA2
MNK1/2
RSK2
CREB
Elk-1
Sap1a
c-Myc
SRF
c-fos
DNA
Proliferation, differentiation
PPP3C
GSTπ
HSP72
Evi1
FLNA
JIP3
ARRB
CrkII
GLK
HGK
HPK1
Tpl2/Cot
MEKK1
MLK3
Cdc42/Rac
PAK1/2
MEKK2/3
MKK4
MKK7
JIP1/2
JNK
NFAT-2
NFAT-4
c-JUN
JunD
ATF-2
Elk-1
p53
p53 signaling pathway
Proliferation, differentiation, inflammation
Apoptosis
Cell cycle
Serum, cytotoxic drugs, irradiation, heatshock, reactive oxygen species, lipopolysaccharide, and other stress
LZK
MUK
MLTK
PP2CA
AKT
PTP
MKP
CASP
MST1/2
TNF
IL1
FASL
TGFB
LPS
TNFR
IL1R
FAS
TGFBR
CD14
TRAF2
GCK
ASK2
ASK1
DAXX
TAB1
TAB2
ECSIT
TAK1
PP2CB
TRAF6
MKK3
MKK6
p38
PP5
Sap1a
GADD153
MAX
MEF2C
PRAK
MAPKAPK
MSK1/2
HSP27
CREB
Cdc25B
DNA damage
GADD45
MEKK4
TAO1/2
NLK
Wnt signaling pathway
Serum, EGF, reactive oxygen species, or Src tyrosinkinase downstream
MEK5
ERK5
Nur77
MAPKKKK
MAPKKK
MAPKK
MAPK
Transcription factor
04010 7/24/13
(c) Kanehisa Laboratories

Other Pathway Databases

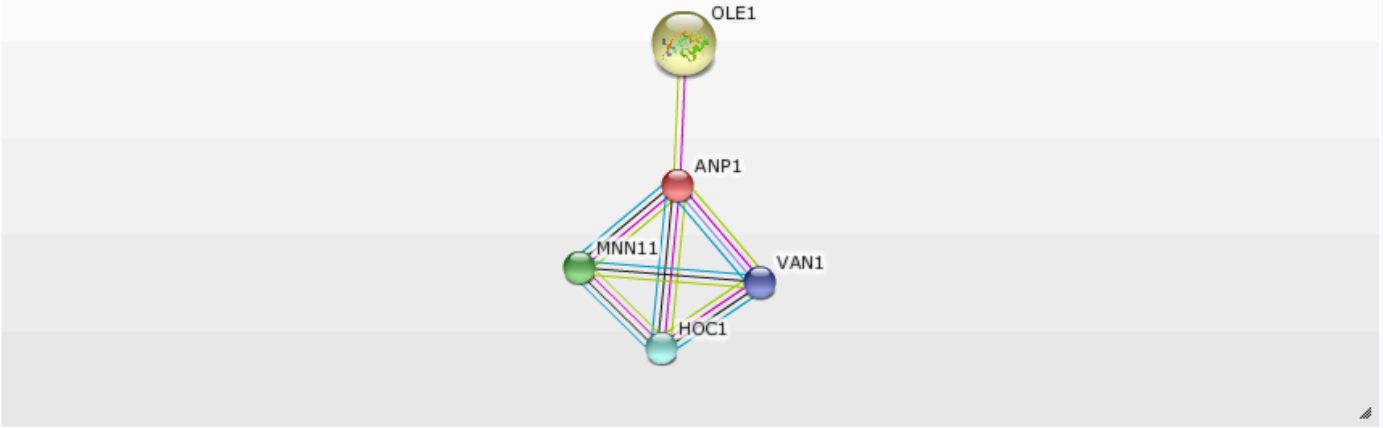
http://www.reactome.org/

http://www.pathwaycommons.org/

STRING


Version: 10.5
LOGIN
REGISTER
STRING
http://string-db.org/
Search
Download
Help
My Data
Welcome to STRING
Protein-Protein Interaction Networks
ORGANISMS
2031
PROTEINS
9.6 mio
INTERACTIONS
1380 mio
SEARCH
© STRING CONSORTIUM 2017
SIB - Swiss Institute of Bioinformatics
CPR - NNF Center for Protein Research
EMBL - European Molecular Biology Laboratory
ABOUT
Content
References
Contributors
Statistics
INFO
Scores
Use scenarios
FAQs
Cookies/Privacy
ACCESS
Versions
APIs
Licensing
Usage
CREDITS
Funding
Datasources
Partners
Software

Save Layout Clustering Enrichment Options



This is the **evidence view**. Different line colors represent the types of evidence for the association.

(requires Flash player 10 or better)

Your Input:

- ANP1 Mannan polymerase II complex ANP1 subunit (M-Pol II subunit ANP1) (Aminonitrophenyl propanediol resistance protein); Involved in the organization of the secretory pathway. Required to maintain a functional Golgi apparatus (500 aa)
- OLE1 Acyl-CoA desaturase 1 (EC 1.14.19.1) (Stearoyl-CoA desaturase 1) (Fatty acid desaturase 1); Utilizes O(2) and electrons from the reduced cytochrome b(5) domain to catalyze the insertion of a double bond into a spectrum of fatty acyl-CoA substrates (Probable) (510 aa)
- MNN11 Probable alpha-1,6-mannosyltransferase MNN11 (EC 2.4.1.-) (Mannan polymerase II complex MNN11 subunit) (M-Pol II subunit MNN11); Required for synthesis of full length mannan chains (422 aa)
- HOC1 Putative glycosyltransferase HOC1 (EC 2.4.-.-) (Mannan polymerase II complex HOC1 subunit) (M-Pol II subunit Hoc1p); The M-Pol II complex possesses alpha-1,6- mannosyltransferase activity and is probably involved in the elongation of the mannan backbone of N-linked glycans on cell wall and periplasmic proteins (396 aa)
- VAN1 Mannan polymerase I complex VAN1 subunit (M-pol I subunit VAN1) (Vanadate resistance protein); Involved in regulation of the phosphorylation of a number of proteins, some of which appear to be important in cell growth control (535 aa)

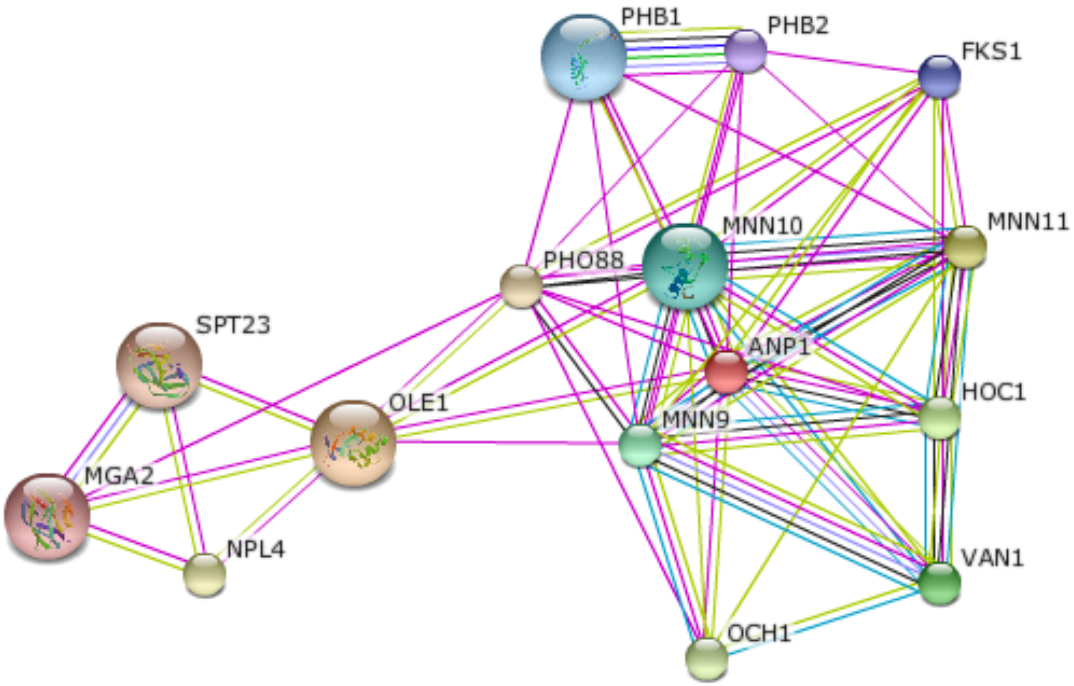
(Saccharomyces cerevisiae)

Neighborhood
Gene Fusion
Cooccurrence
Coexpression
Experiments
Databases
Textmining
[Homology]
Score

Predicted Functional Partners:

Views:

Neighborhood Fusion Occurence Coexpression Experiments Database Textmining Summary Network



This is the **evidence view**. Different line colors represent the types of evidence for the association.

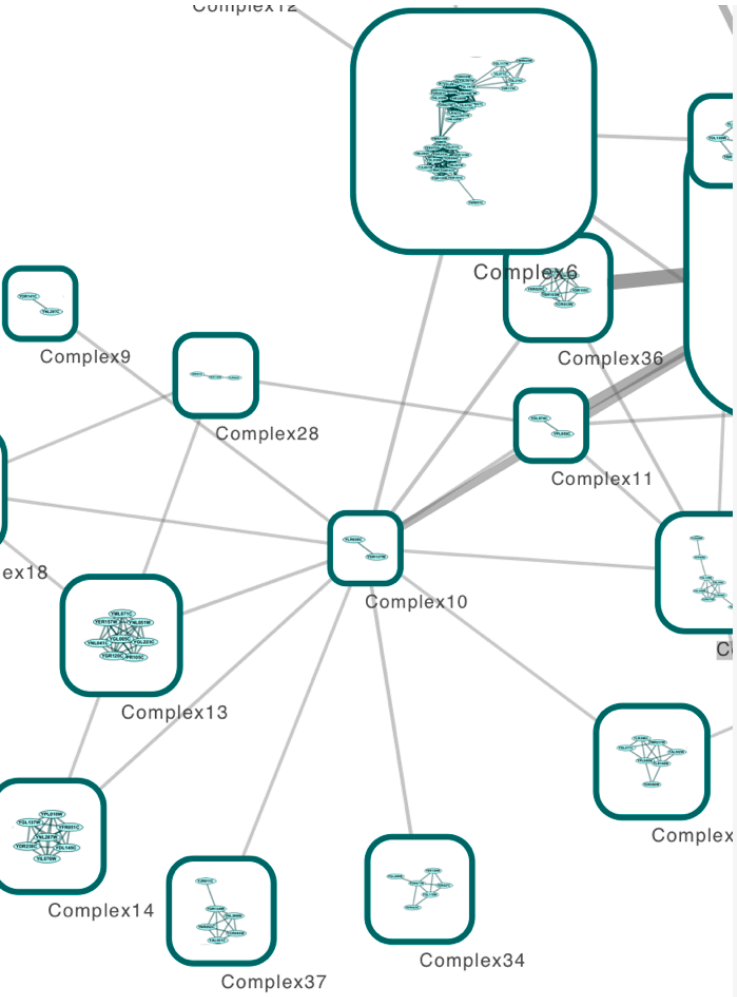
confidence evidence actions interactive advanced more less save

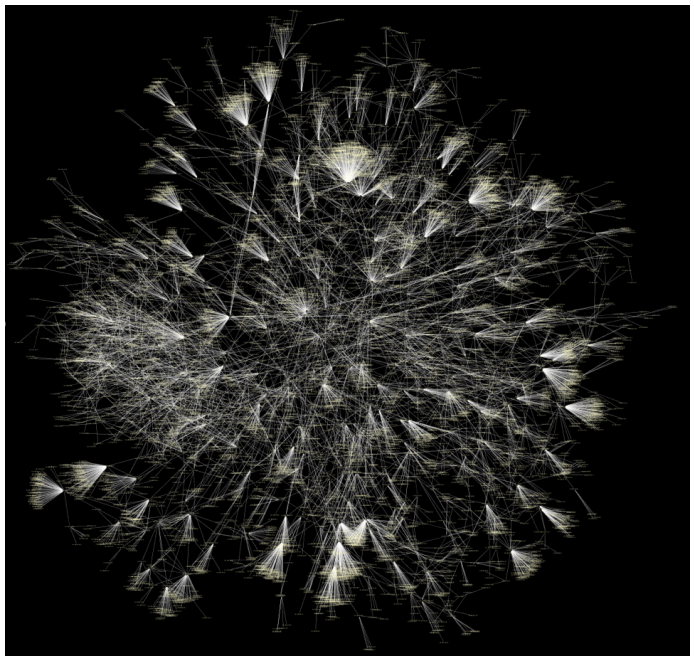
(requires Flash player 10 or better)

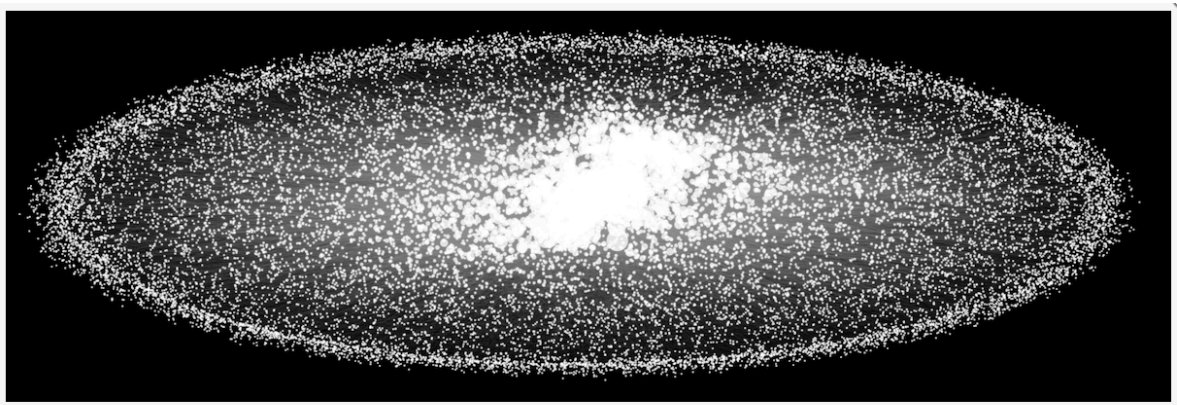
Cytoscape


Cytoscape
Search...
Go
Home
Introduction
Download
Apps
Documentation
Community
Report a Bug
Getting Help
Cytoscape
Network Data Integration, Analysis, and Visualization in a Box
Cytoscape is an open source software platform for visualizing complex networks and integrating these with any type of attribute data. A lot of Apps are available for various kinds of problem domains, including bioinformatics, social network analysis, and semantic web.
Download Cytoscape
Welcome Letter
Release Notes
Sample Visualizations
Learning Cytoscape
http://www.cytoscape.org/
Introduction to Cytoscape
Tutorial for 3.x
Tutorial for 2.x
Developers Tutorial

Small and Large networks







Human Interactome Visualization: Cytoscape can handle large network data sets

All Apps

Categories

- network generation
- online data import
- graph analysis
- data visualization
- integrated analysis
- utility
- clustering
- ontology analysis
- enrichment analysis
- scripting

more »

Featured Apps

Get Started with the App Store »

PSICQUICUniversalClient
PSICQUIC Web Service Client for importing interactions from public

BiNGO 3.0+
Calculates overrepresented GO terms in the network and display

ClueGO 3.0+
Creates and visualizes a functionally grouped network of

DynNetwork 3.0+
Visualize dynamic networks in Cytoscape 3.0

Top Voted Apps

ClueGO 3.0+
Creates and visualizes a functionally grouped network of

GeneMANIA 3.0+
Imports interaction networks from public databases from a list of

CytoKegg 3.0+
Identify Kegg pathways associated to specific

PathExplorer 3.0+
Finds paths, filters them based on node and edge attributes and

Top Downloaded Apps

ClueGO 3.0+
Creates and visualizes a functionally grouped network of

GeneMANIA 3.0+
Imports interaction networks from public databases from a list of

MCODE 3.0+
Clusters a given network based on topology to find densely

jActiveModules 3.0+
Finds clusters where member nodes show significant changes

BiNGO

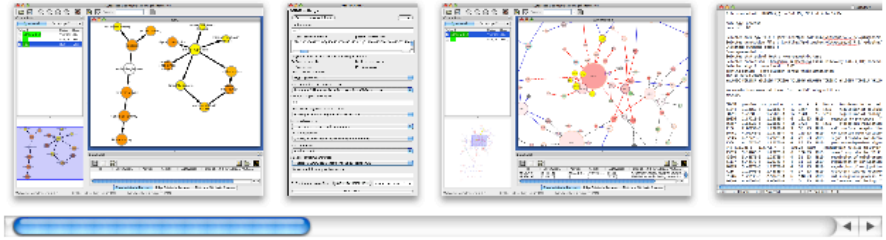
Calculates overrepresented GO terms in the network and display them as a network of significant GO terms.

★★★★☆ (21) 2639 downloads 3.0+

BiNGO

Details | Release History

Categories: enrichment analysis, GO annotation, ontology analysis



BiNGO is a tool to determine which Gene Ontology (GO) categories are statistically overrepresented in a set of genes or a subgraph of a biological network. BiNGO maps the predominant functional themes of a given gene set on the GO hierarchy, and outputs this mapping as a Cytoscape graph. Gene sets can either be selected or computed from a Cytoscape network (as subgraphs) or compiled from sources other than Cytoscape (e.g. a list of genes that are significantly upregulated in a microarray experiment). The main advantage of BiNGO over other GO tools is the fact that it can be used directly and interactively on molecular interaction graphs. Another plus is that BiNGO takes full advantage of Cytoscape's versatile visualization environment. This allows you to produce customized high-quality figures. Features include :

1. Assessing over-representation or under-representation of GO categories
2. Graph or gene list input
3. batch mode: analyze several clusters simultaneously using same settings
4. Different GO and GOSlim ontologies
5. Wide range of organisms
6. Evidence code filtering
7. Hypergeometric or binomial test for over-representation
8. Multiple testing correction using Bonferroni (FWER) or Benjamini&Hochberg (FDR) correction
9. Interactive visualization of results mapped on the GO hierarchy
10. Extensive results in tab-delimited text file format
11. Make and use custom annotations, ontologies and reference sets
12. Open source

If you use BiNGO in your research, please cite:

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. ***Bioinformatics*** *21, 3448-3449.* (PubMed)

BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 1 2014, pages 135–136
doi:10.1093/bioinformatics/btt598

Systems biology

Advance Access publication October 21, 2013

BEReX: Biomedical Entity-Relationship eXplorer

Minji Jeon[1], Sunwon Lee[2], Kyubum Lee[2], Aik-Choon Tan[3] and Jaewoo Kang[1,2,*]

[1]Interdisciplinary Graduate Program in Bioinformatics, [2]Department of Computer Science and Engineering, Korea University, Seoul, Korea and [3]Department of Medicine/Medical Oncology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Associate Editor: Mario Albrecht

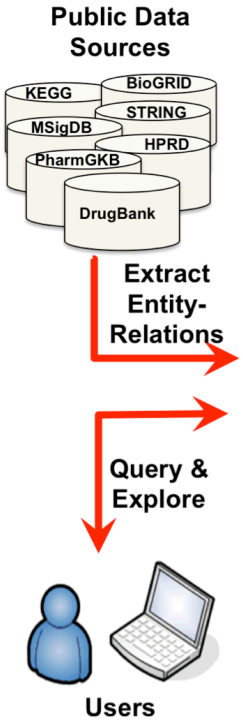
ABSTRACT

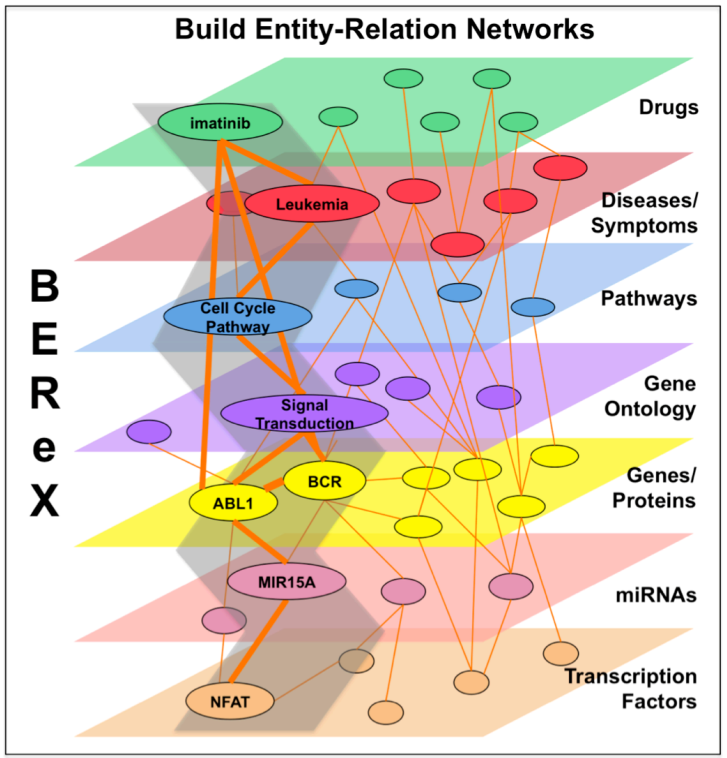
Summary: Biomedical Entity-Relationship eXplorer (BEReX) is a new biomedical knowledge integration, search and exploration tool. BEReX integrates eight popular databases (STRING, DrugBank, KEGG, PhamGKB, BioGRID, GO, HPRD and MSigDB) and delineates an integrated network by combining the information available from these databases. Users search the integrated network by entering key words, and BEReX returns a sub-network matching the key words. The resulting graph can be explored interactively. BEReX allows users to find the shortest paths between two remote nodes, find the most relevant drugs, diseases, pathways and so on related to the current network, expand the network by particular types of entities and relations and modify the network by removing or adding selected nodes. BEReX is implemented as a standalone Java application.

Availability and implementation: BEReX and a detailed user guide are available for download at our project Web site (http://infos.korea.ac.kr/berex).

Contact: kangj@korea.ac.kr

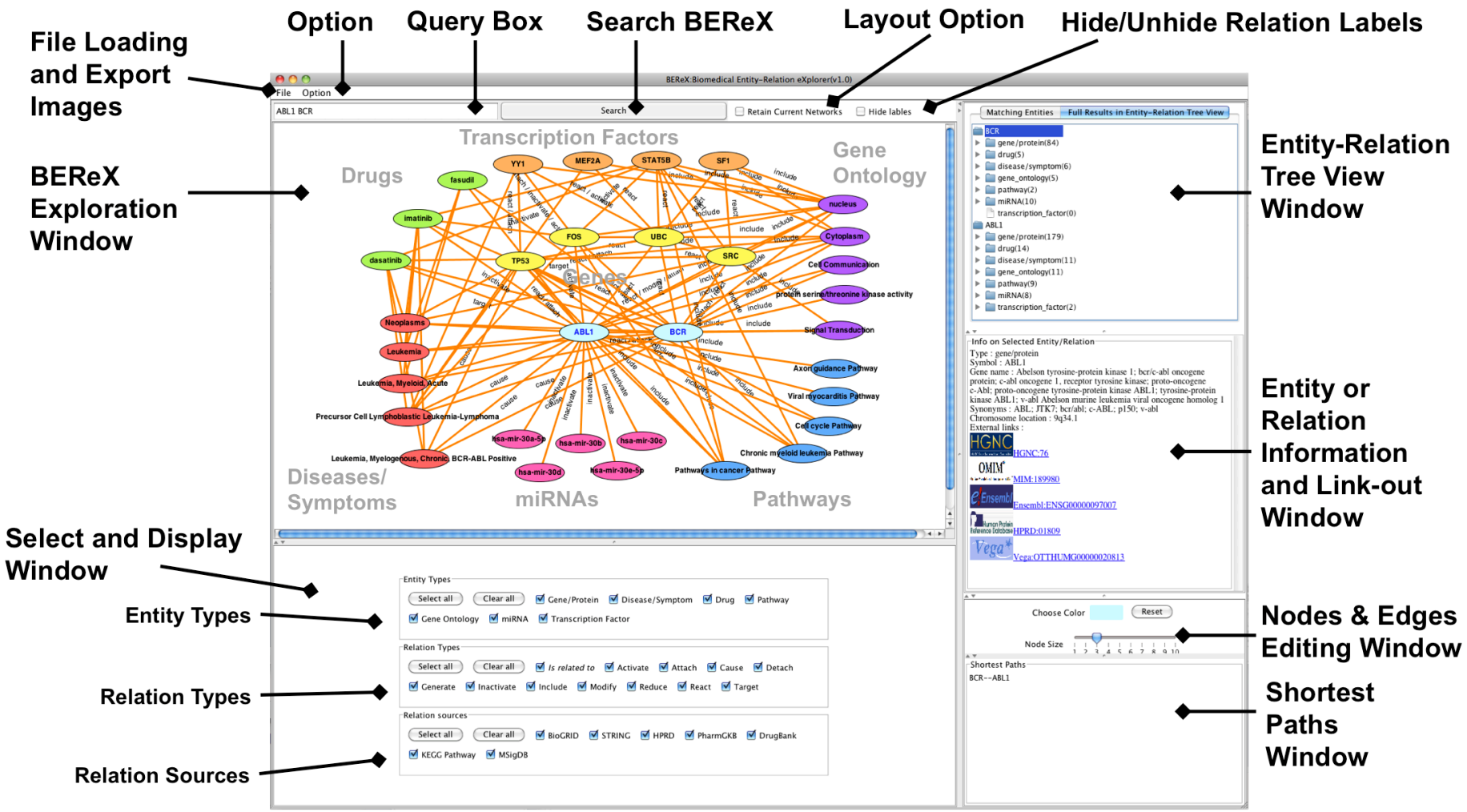
Supplementary Information: Supplementary methods and user guide are available at *Bioinformatics* online.





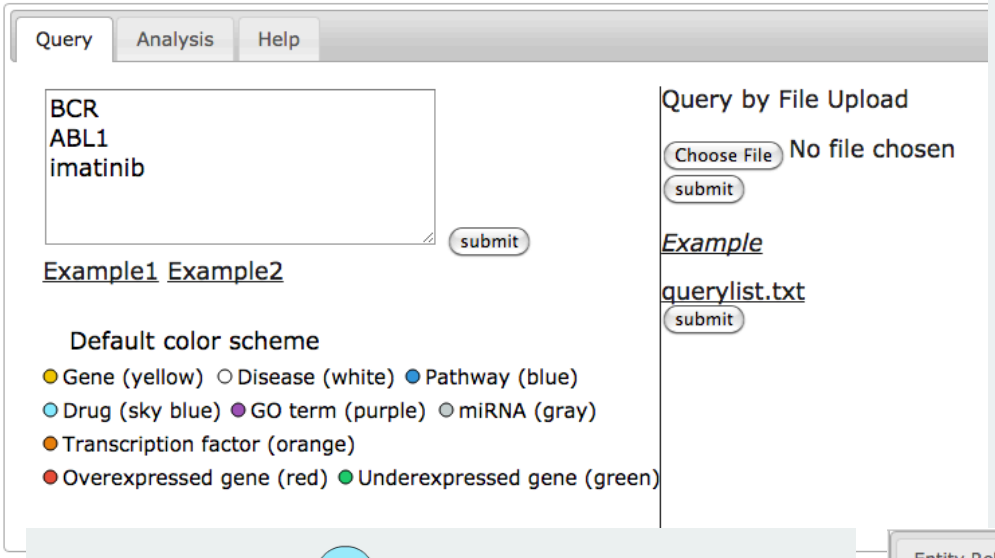
(Collaboration with Dr. Jaewoo Kang, Korea University)

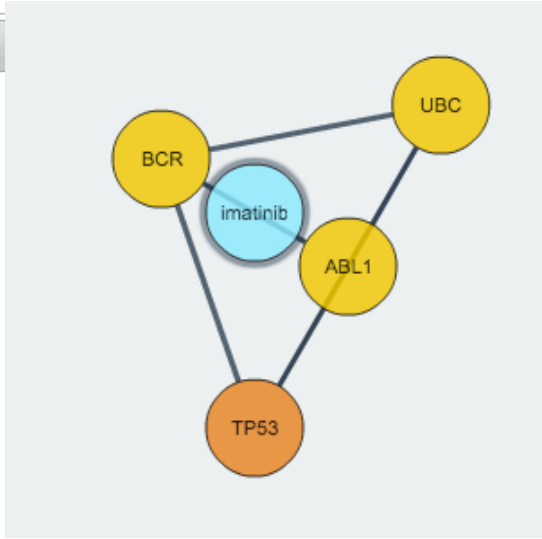
BEReX: Biomedical Entity-Relationship eXplorer

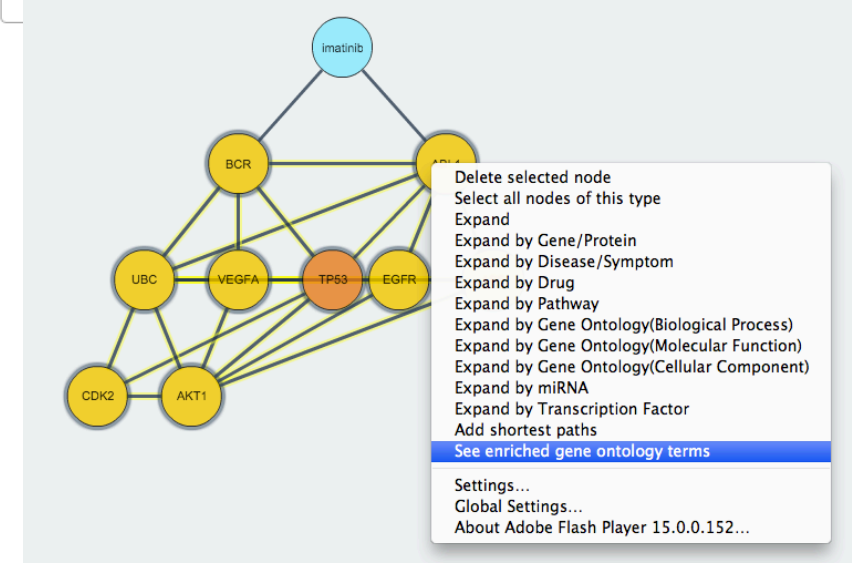


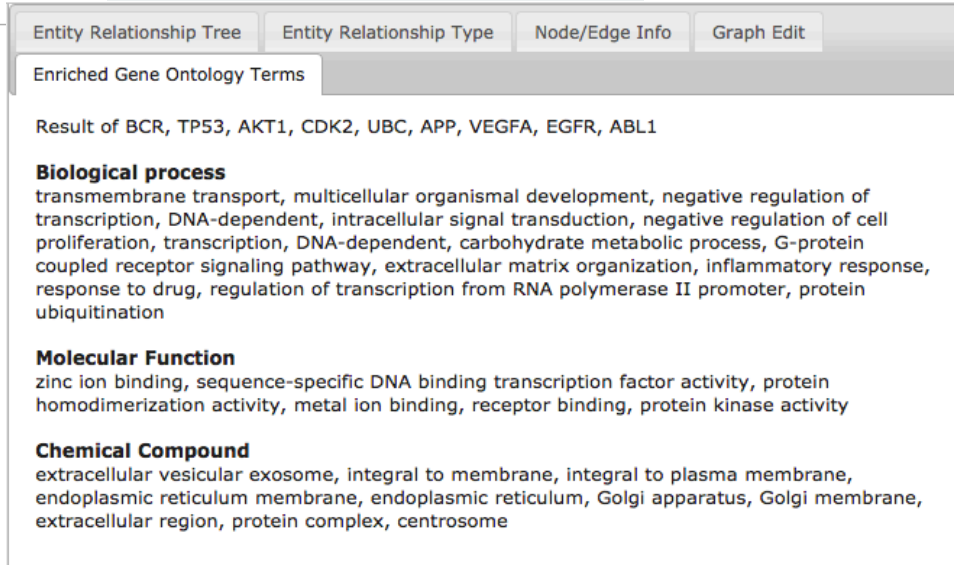
(Collaboration with Dr. Jaewoo Kang, Korea University)

*i*BEReX (http://iberex.korea.ac.kr/)









(Collaboration with Dr. Jaewoo Kang, Korea University)

Take Home Message

- Motivation to do gene list enrichment analysis.
- Choose the right statistics for the enrichment analysis – depending on what you want to ask.
- Use p-values carefully.
- Many enrichment tools are available, test multiple tools.
- Use interactive visualization tools to present your enriched lists of genes.