

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

¹Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada.

²Department of Molecular Genetics, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. ³Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. ⁴Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

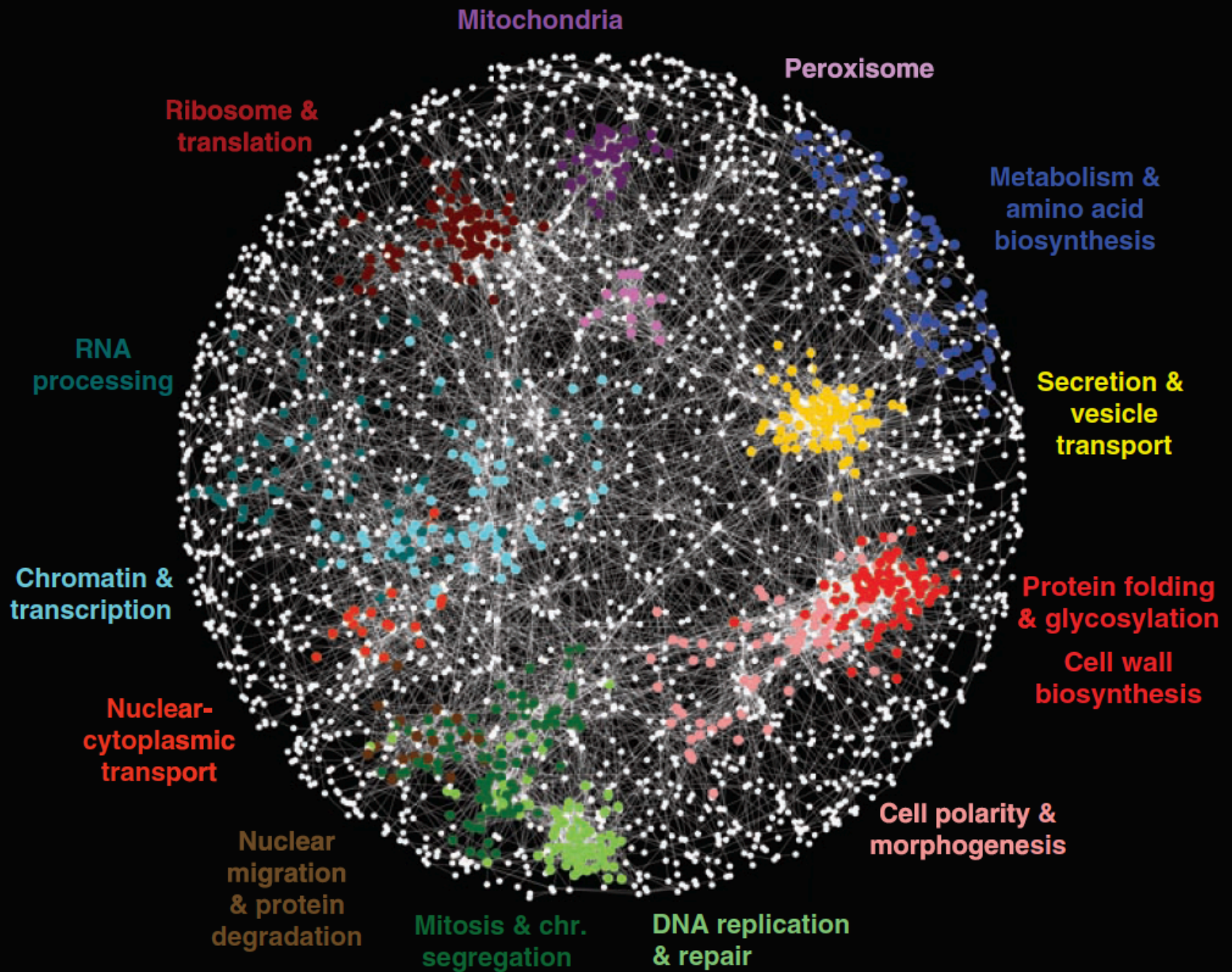
⁵Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada. ⁶Department of Biochemistry, Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304, USA. ⁷Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

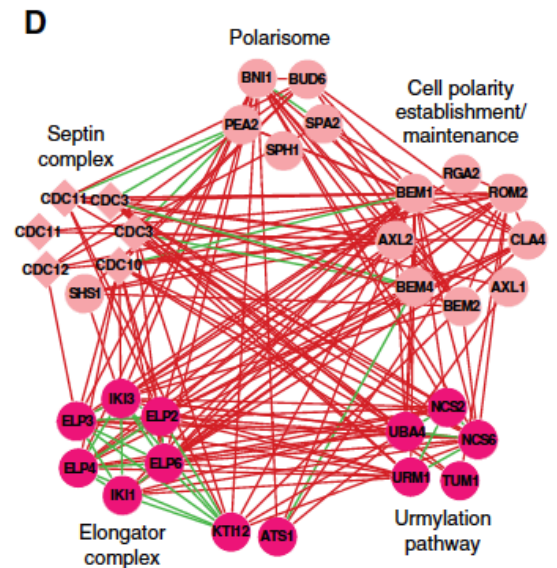
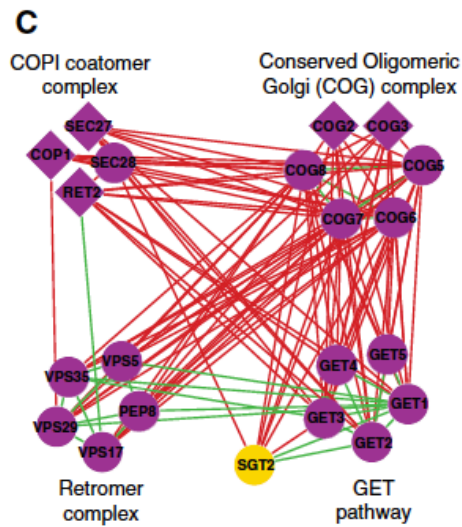
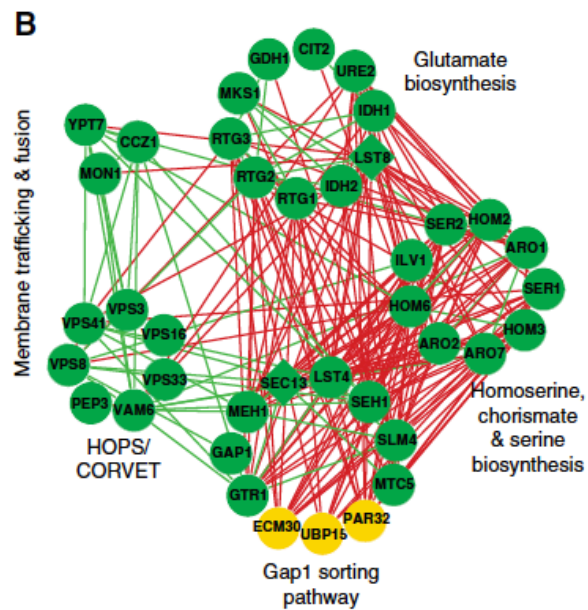
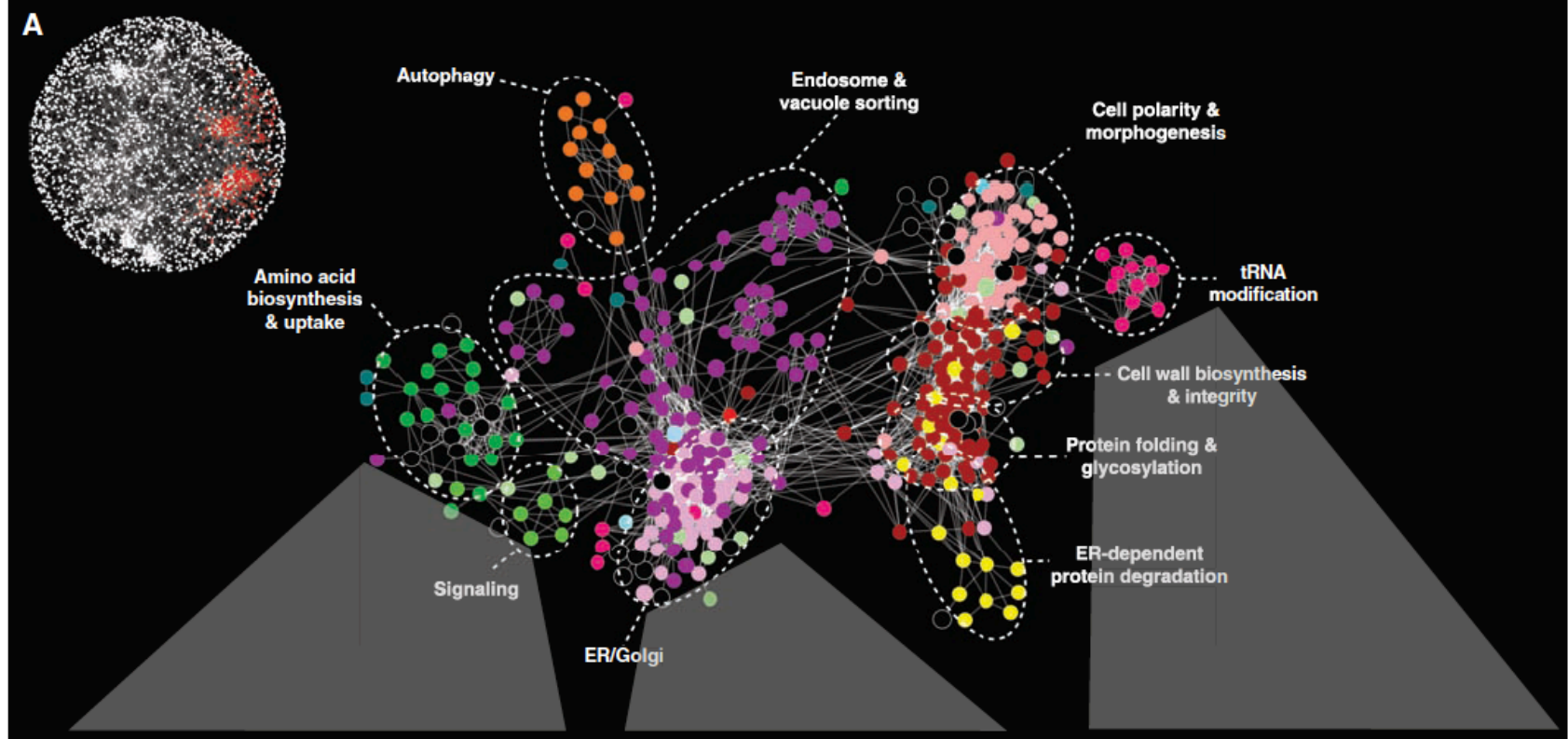
⁸Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

⁹Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada.

¹⁰Department of Pharmacy, University of Toronto, Toronto, Ontario M5S 3E1, Canada. ¹¹Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA. ¹²S&P Robotics, Inc., 1181 Finch Avenue West, North York, Ontario M3J 2V8, Canada. ¹³Institute of Biochemistry, Biological Research Center, H-6701 Szeged, Hungary.

¹⁴Department of Computer Science, Lewis-Sigler Institute for Integrative Genomics, Carl Icahn Laboratory, Princeton University, Princeton, NJ 08544, USA. ¹⁵Biology Department, McGill University, Montreal, Quebec H3A 1B1, Canada.



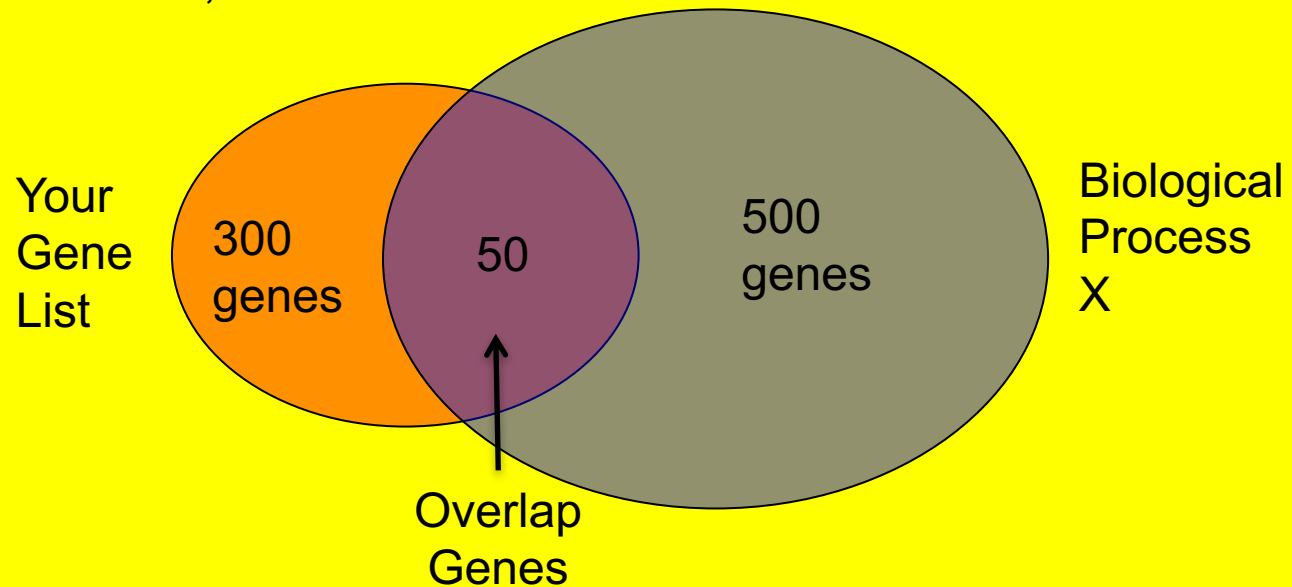


Why Do Enrichment Analysis?

- 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

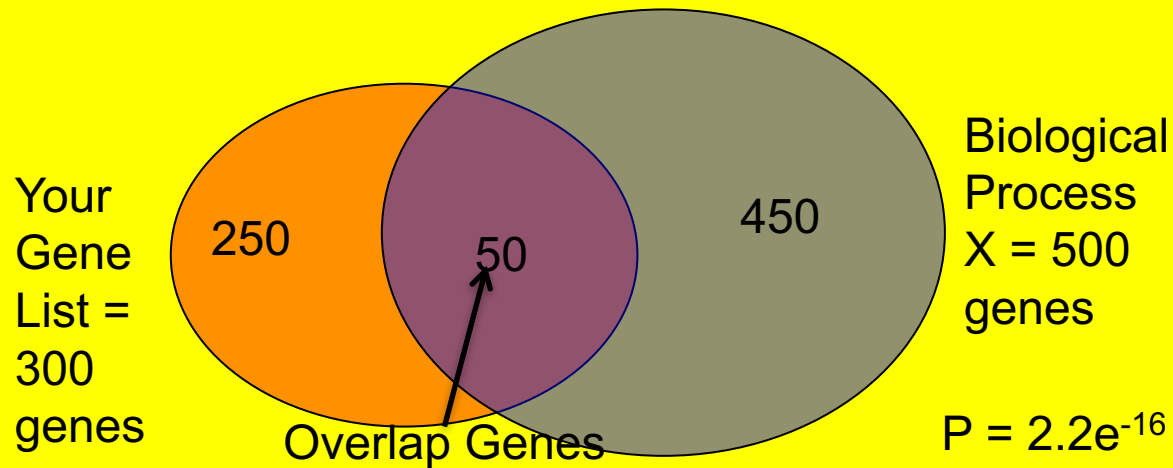


Different Tests for Enrichment

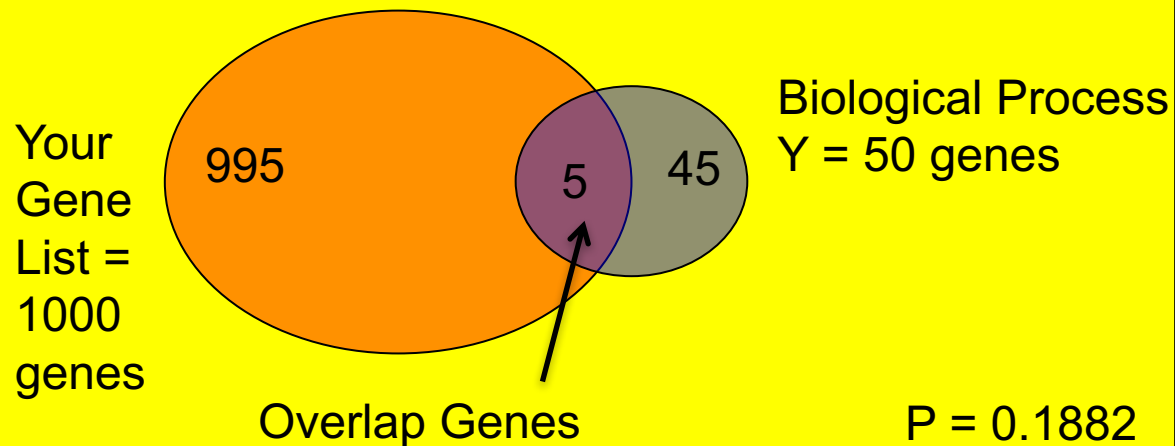
- Fisher's exact
- Chi-square
- Binomial
- Hypergeometric
- Kolmogorov-Smirnov *(This is implemented in GSEA)*
- Permutation
- ...

Statistical Test for Enrichment

Assume Total Genes in Genome = 20,000



Assume Total Genes in Genome = 20,000



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)

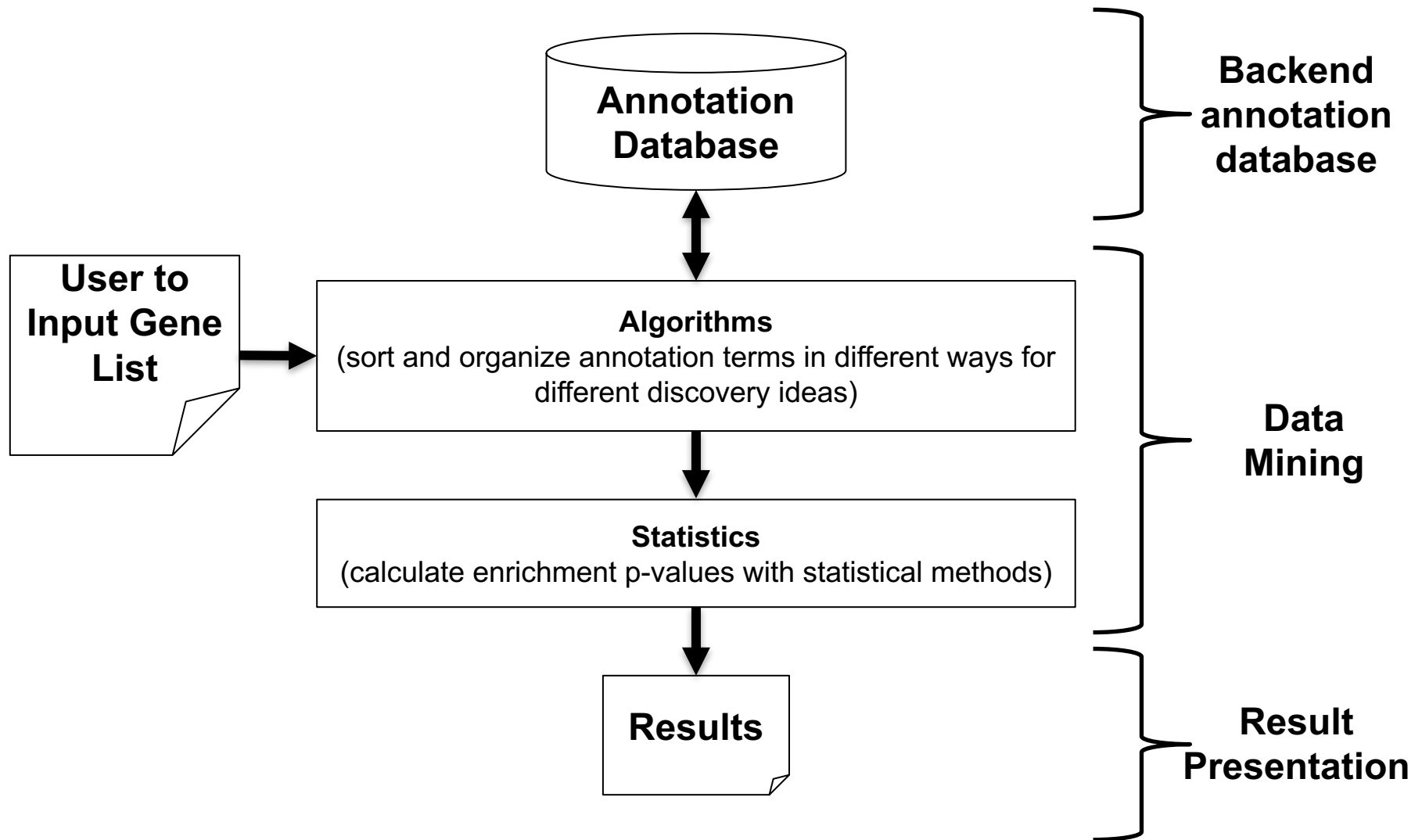
Input for Enrichment Test

- 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?

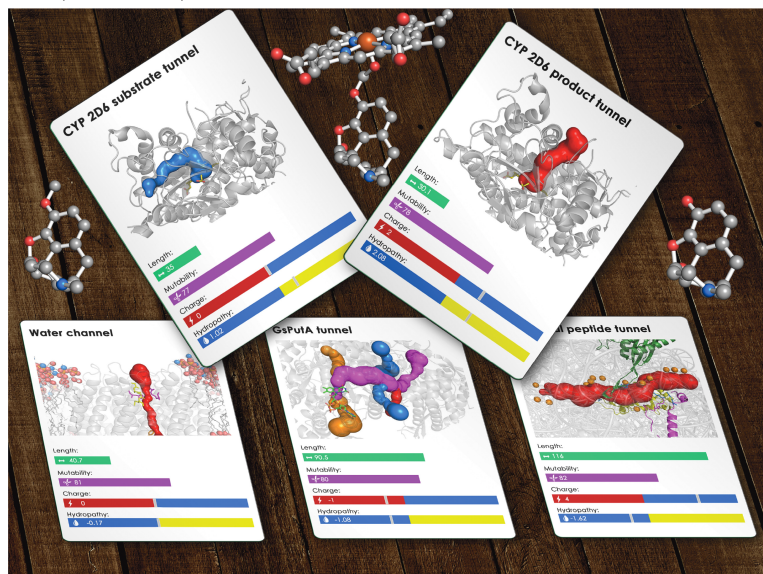
181 databases,
Online: 1737 databases
<http://www.oxfordjournals.org/nar/database/c/>

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>



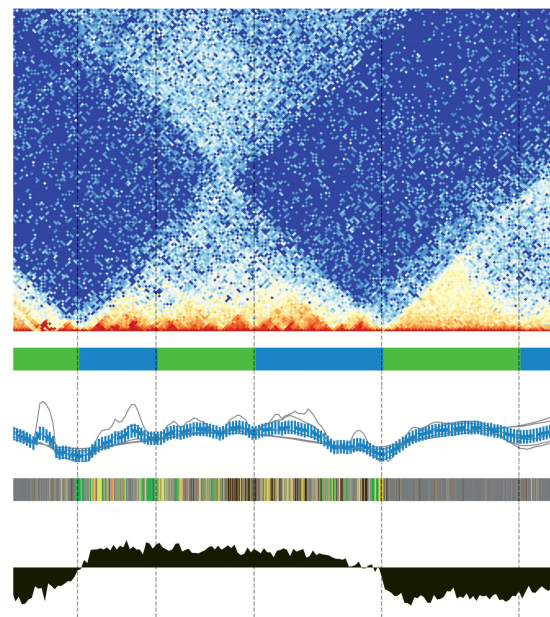
85 web servers

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



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

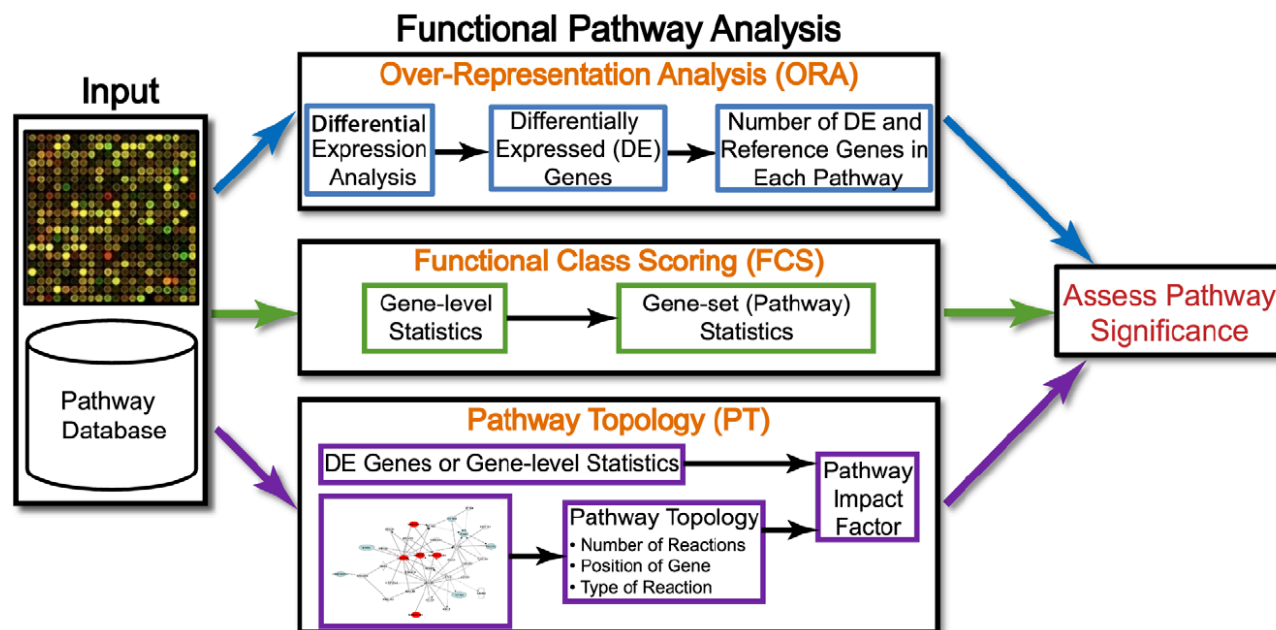


Table 1. Examples of pathway analysis tools in each generation.

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://lama.mshri.on.ca/funcassociate/)	[6]
GOToolBox	Web (http://genome.crg.es/GOToolBox/)	[10]
GeneMerge	Standalone, Web (http://genemerge.cbc.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]

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; hypergeometric; 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-square	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	Hypergeometric	Class I
JProGO	2006	Fisher's exact; Kolmogorov-Smirnov test; student's <i>t</i> -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; <i>t</i> -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	<i>t</i> -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
POSO	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
PaS	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 <i>P</i> -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, GToolBox, 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 <i>P</i> -value calculation.	Suitable for pair-wise 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	<i>t</i> -Test permutation Z-score	FatiScan, ADGO, ermineJ, PAGE, iGA, GO-Mapper, GDist, 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 <i>P</i> -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

**DAVID Bioinformatics Resources 6.8**
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Home | **Start Analysis** | **Shortcut to DAVID Tools** | Technical Center | Downloads & APIs | Term of Service | Why DAVID? | About Us

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

Shortcut to DAVID Tools

Functional Annotation
Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and more

Gene Functional Classification
Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological context captured by high throughput technologies. [More](#)

Gene ID Conversion
Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. [More](#)

Gene Name Batch Viewer
Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. [More](#)

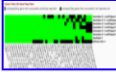
Recommending: A [paper](#) published in *Nature Protocols* describes step-by-step procedure to use DAVID!

Welcome to DAVID 6.8

2003 - 2017

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 comprises a full Knowledgebase update to the sixth version of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

☒ Identify enriched biological themes, particularly GO terms
☒ Discover enriched functional-related gene groups
☒ Cluster redundant annotation terms
☒ Visualize genes on BioCarta & KEGG pathway maps
☒ Display related many-genes-to-many-terms on 2-D view.
☒ Search for other functionally related genes not in the list
☒ List interacting proteins
☒ Explore gene names in batch
☒ Link gene-disease associations
☒ Highlight protein functional domains and motifs
☒ Redirect to related literatures
☒ Convert gene identifiers from one type to another.
☒ And more

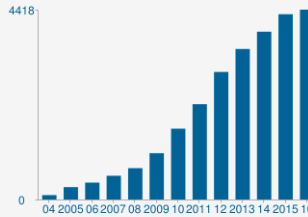


What's Important in DAVID?

- [Cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)
- [Novel Classification Algorithms](#)
- [Pre-built Affymetrix and Illumina backgrounds](#)
- [User's customized gene background](#)
- [Enhanced calculating speed](#)

Statistics of DAVID

DAVID Bioinformatic Resources Citations



Year	Citations
04	100
2005	200
06	300
2007	400
08	500
2009	600
10	700
2011	800
12	900
2013	1000
14	1100
2015	1200
16	1300

- [> 26,000 Citations](#)
- Average Daily Usage: ~2,600 gene lists/sublists from ~800 unique researchers.
- Average Annual Usage: ~1,000,000 gene lists/sublists from >5,000 research institutes world-wide

Please cite [Nature Protocols 2009; 4\(1\):44](#) & [Nucleic Acids Res. 2009; 37\(1\):1](#) within any publication that makes use of any methods inspired by DAVID.



[Term of Service](#) | [Contact Us](#) | [Site Map](#)

*** 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

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

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			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			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
☐	PIRSF001875:alpha-defensin	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](#)

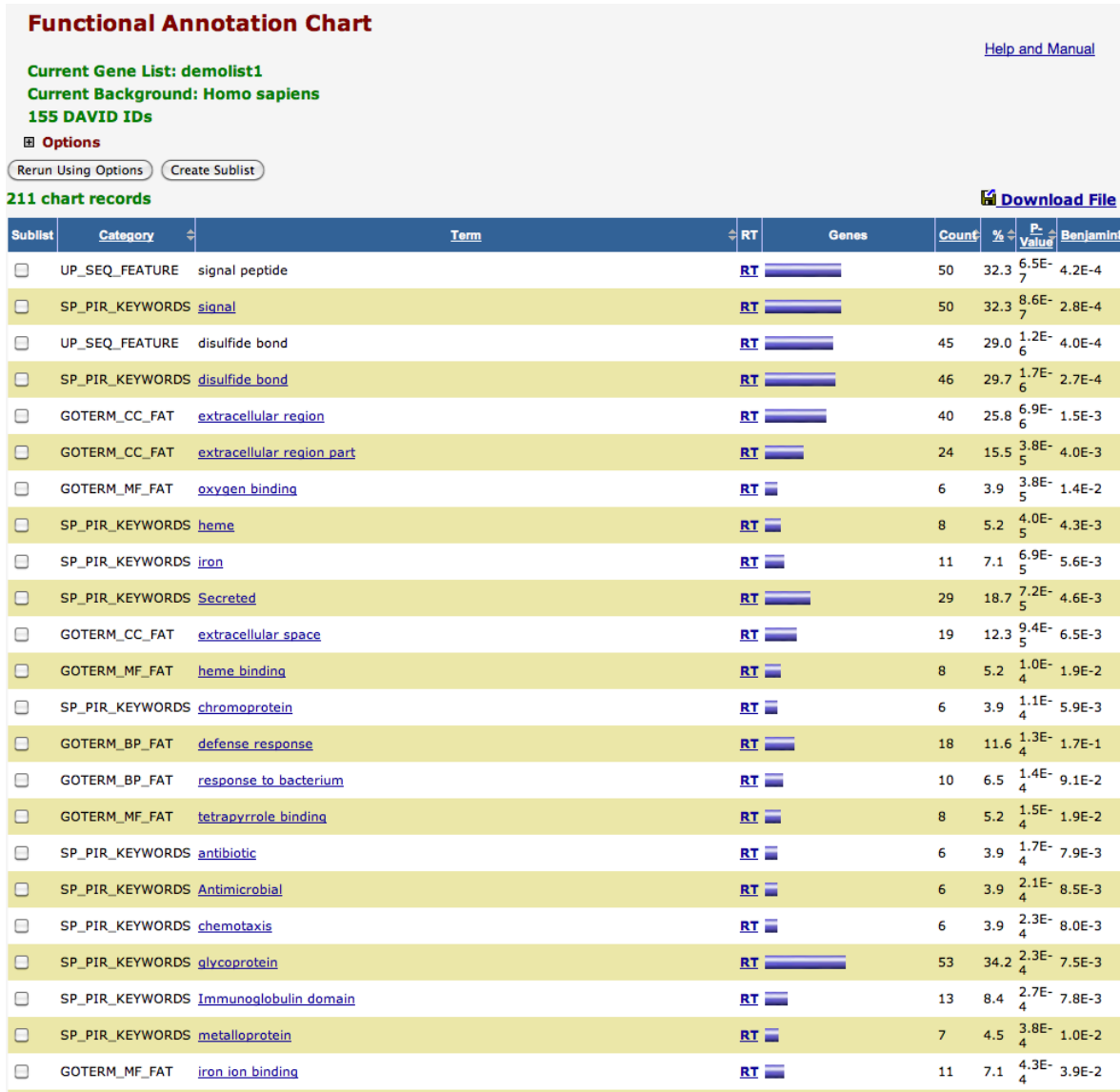
Similarity Score: ☒ Very High (0.75-1) ☐ High (0.5-0.75) ☐ Moderate (0.25-0.5) ☐ Low (<0.25)

#	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

■ corresponding gene-term association positively reported ■ corresponding gene-term association not reported yet

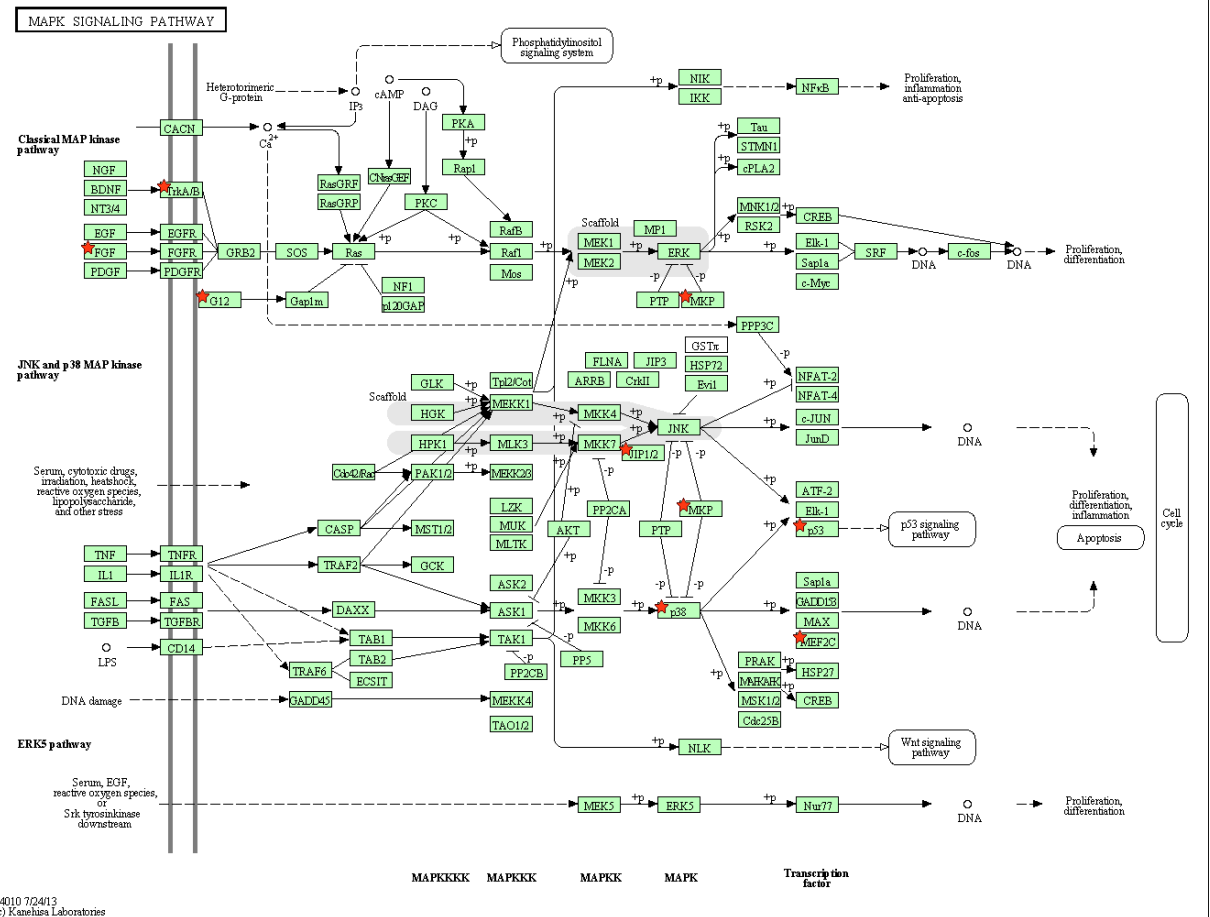


Functional Annotation Chart



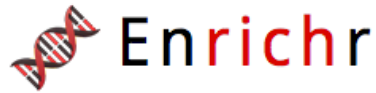
Linking Enriched Genes to Pathways (KEGG only)

☐ 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/>



[Login](#) | [Register](#)

14,777,355 lists analyzed

247,533 terms

133 libraries

Analyze

[What's New?](#)

[Libraries](#)

[Find a Gene](#)

[About](#)

[Help](#)

Input data

Choose an input file to upload. Either in BED format or a list of genes. For a quantitative set, add a comma and the level of membership of that gene. The membership level is a number between 0.0 and 1.0 to represent a weight for each gene, where the weight of 0.0 will completely discard the gene from the enrichment analysis and the weight of 1.0 is the maximum.

Try an example [BED file](#).

No file chosen

Or paste in a list of gene symbols optionally followed by a comma and levels of membership. Try two examples: [crisp set example](#), [fuzzy set example](#)

0 gene(s) entered

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

☐ **Contribute**

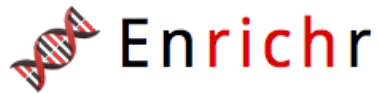
Please acknowledge Enrichr in your publications by citing the following references:

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).

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016; gkw377.

Enrichr

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



[Login](#) | [Register](#)

14,777,355 lists analyzed

247,533 terms

133 libraries

Analyze

[What's New?](#)

[Libraries](#)

[Find a Gene](#)

[About](#)

[Help](#)

Input data

Choose an input file to upload. Either in BED format or a list of genes. For a quantitative set, add a comma and the level of membership of that gene. The membership level is a number between 0.0 and 1.0 to represent a weight for each gene, where the weight of 0.0 will completely discard the gene from the enrichment analysis and the weight of 1.0 is the maximum.

Try an example [BED file](#).

No file chosen

Or paste in a list of gene symbols optionally followed by a comma and levels of membership. Try two examples: [crisp set example](#), [fuzzy set example](#)

```
Nsun3
Polrmt
Nlrp1
Sfxn5
Zc3h12c
Slc25a39
Arsg
Defb29
Ndufb6
Zfand1
```

375 gene(s) entered

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

☐ **Contribute**

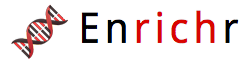
Please acknowledge Enrichr in your publications by citing the following references:

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).

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016;gkw377.

Enrichr

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



[Login](#) | [Register](#)

Transcription Pathways Ontologies Disease/Drugs Cell Types Misc Legacy Crowd

Description Sample gene list (375 genes)



ChEA 2016



E2F1_18555785_ChIP-Seq_MESCs_Mouse
SRF_Z1415370_ChIP-Seq_H1_Mouse
JARID1A_20064375_ChIP-Seq_MESCs_Mouse
NELFA_Z0434984_ChIP-Seq_ESCs_Mouse
PPARA_Z2152653_ChIP-Seq_LIVER_Mouse

ENCODE and ChEA Consensus TFs from



ERG_CHEA
NR2C2_ENCODE
GABPA_ENCODE
HNF4A_ENCODE
ZMIZ1_ENCODE

ARCHS4 TFs Coexp



NPAS1_human_tf_ARCHS4_coexpression
ZNF17_human_tf_ARCHS4_coexpression
ZNF7808_human_tf_ARCHS4_coexpression
GABPB1_human_tf_ARCHS4_coexpression
ZBTB6_human_tf_ARCHS4_coexpression

TF Perturbations Followed by Expression



KLF3_KO_MOUSE_GSE43806_CREEDSID_GEN
IRF9_OE_HUMAN_GSE50002_CREEDSID_GEN
GLIS2_KO_MOUSE_GSE6113_CREEDSID_GEN
IRF9_OE_HUMAN_GSE50002_CREEDSID_GEN
TWIST1_OE_MOUSE_GSE50002_CREEDSID_G

Enrichr Submissions TF-Gene Cooccurrence



ZRSR1
ZNF830
MYNN
NUP12
TRIM23

TRANSFAC and JASPAR PWMs



CEBPE (human)
SOX2 (human)
FOXA1 (mouse)
NFATC3 (human)
FOXF2 (human)

Epigenomics Roadmap HM ChIP-seq



H3K9me3 Fetal Heart
H3K36me3 Muscle Satellite Cultured Cells
H3K27ac CD4 Naive Primary Cells
H3K27ac H9
H3K79me2 H9

TargetScan microRNA 2017



hsa-miR-1278
mmu-miR-704
mmu-miR-5099
hsa-miR-3144-3p
mmu-miR-707

miRTarBase 2017



mmu-miR-140-5p
hsa-miR-186-3p
hsa-miR-450a-5p
hsa-miR-190a-3p
mmu-miR-3086-5p

ENCODE TF ChIP-seq 2015



POLR2A_kidney_mm9
POLR2AphosphoS5_Panc1_hg19
NR2C2_GM12878_hg19
SPT_A549_hg19
MYC_MEL_cell_line_mm9

TF-LOF Expression from GEO



glis2_17618285_kidney_lof_mouse_gpi2897
creb1_22108299_heart_left_ventricle_lof_mo
stat3_18500982_mesc_gof_mouse_gpi83_gdi
wt1_00000000_mouse_embryonic_fibroblast
hnf4a_16714333_e18dot5_liver_lof_mouse_g

ENCODE Histone Modifications 2015



H3K27ac_liver_mm9
H3ac_myocyte_mm9
H3K27ac_CH12.LX_mm9
H3K9ac_skeletal_muscle_myoblast_hg19
H3K27ac_cerebellum_mm9

Transcription Factor PPIs



PPARGC1A
NR4A1
ATF1
CREM
GLI1

Genome Browser PWMs



GKCGCNNNNNNTGATG_UNKNOWN
VSPXR_Q2
VSAHR_Q5
VSGNCF_Q1
CAGNYGKNAAA_UNKNOWN

Enrichr

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



Enrichr

[Login](#) | [Register](#)

Transcription **Pathways** Ontologies Disease/Drugs Cell Types Misc Legacy Crowd

Description Sample gene list (375 genes)



KEGG 2016



Metabolic pathways_Homo sapiens_hsa01111
Glyoxylate and dicarboxylate metabolism_H
Tyrosine metabolism_Homo sapiens_hsa0003
Carbon metabolism_Homo sapiens_hsa01202
Beta-alanine metabolism_Homo sapiens_hsa

WikiPathways 2016



Mitochondrial Gene Expression_Mus muscu
Mitochondrial Gene Expression_Homo sapie
Amino Acid metabolism_Mus musculus_WP
NAD Biosynthesis II (from tryptophan)_Hom
Methylation Pathways_Homo sapiens_WP70

ARCHS4 Kinases Coexp



TRPM7_human_kinase_ARCHS4_coexpression
AAK1_human_kinase_ARCHS4_coexpression
CDK17_human_kinase_ARCHS4_coexpression
ICK_human_kinase_ARCHS4_coexpression
STK16_human_kinase_ARCHS4_coexpression

Reactome 2016



Metabolism of water-soluble vitamins and c
Glycogen synthesis_Homo sapiens_R-HSA-13
Metabolism_Homo sapiens_R-HSA-1430728
Gamma carboxylation, hypusine formation :
Branched-chain amino acid catabolism_Hom

BioCarta 2016



RNA polymerase III transcription_Homo sap
Visceral Fat Deposits and the Metabolic Syn
SODD/TNFR1 Signaling Pathway_Homo sapi
CARM1 and Regulation of the Estrogen Rece
IL4 Signaling pathway_Homo sapiens_h_IL4F

HumanCyc 2016



glycogen biosynthesis_Homo sapiens_PWW1
tryptophan degradation_Homo sapiens_TY
L-tyrosine degradation_Homo sapiens_PI
Superpathway of tryptophan utilization_Hon
valine degradation_Homo sapiens_VALDEG-I

NCI-Nature 2016



Integrins in angiogenesis_Homo sapiens_26
LPA4-mediated signaling events_Homo sapi
Alpha4 beta1 Integrin signaling events_Hom
Stabilization and expansion of the E-cadheri
Beta3 Integrin cell surface interactions_Hom

Panther 2016



Beta1 adrenergic receptor signaling pathwa
Beta2 adrenergic receptor signaling pathwa
Metabotropic glutamate receptor group II p
Insulin/IGF pathway-mitogen activated prot
Muscarinic acetylcholine receptor 2 and 4 si

BioPlex 2017



NDUFAB1
KERA
GABRG2
SPPL2B
HSPD1

huMAP



ETFA
MRPS30
PRKAR2A
MRPL17
GUSB

PPI Hub Proteins



CDK5
Z31403
PPP2R1A
SKP1
ACTA1

KEA 2015



CDK5
MAPK10
PDPK1
PDK1
ILK2

LINCS L1000 Kinase Perturbations down



GABRP_knockdown_96h_PC3
HRH1_knockdown_96h_HCC515
MKXK2_knockdown_96h_MCF7
PTK7_knockdown_96h_A375
PAK3_knockdown_96h_HCC515

LINCS L1000 Kinase Perturbations up



TNIN_knockdown_96h_MCF7
CXCR2_knockdown_96h_MCF7
BDKR2_knockdown_96h_MCF7
PES_knockdown_96h_MCF7
PIK3CA_knockdown_96h_A375

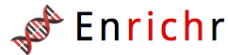
Kinase Perturbations from GEO down



CDK8_knockdown_130_GSE30816
AKT1_activemutant_216_GSE9484
STK11_knockout_278_GSE34866
ATK_knockdown_120_GSE27869
AKT1_knockout_214_GSE39699

Enrichr

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



[Login](#) | [Register](#)

Transcription Pathways Ontologies **Disease/Drugs** Cell Types Misc Legacy Crowd

Description Sample gene list (375 genes) 

LINCS L1000 Chem Pert up

CPC006_PC3_24H-BRD-K35716340-12.0
CPC009_PC3_24H-BRD-K00313977-10.0
CPC002_HA1E_24H-u_99198_maleate-10.0
CPC005_A375_24H-bwb-70c-10.0
CPC002_VCAP_24H-puromycin_hydrochloride

LINCS L1000 Chem Pert down

CPC007_A549_24H-mis-0075347_0001-10.0
CPC001_VCAP_6h-dnc-8_buildingblock04-1
CPC006_A673_6H-BRD-K15563106-177.6
CPC005_VCAP_24H-46_berbamine-10.0
CPC017_MCF7_24H-zk_93426_hydrochloride

DSigDB

ouabain_HL60_DOWN
digoxigenin_HL60_DOWN
strophanthidin_HL60_DOWN
flupentixol_HL60_DOWN
lanatoside C_HL60_DOWN

LINCS L1000 Ligand Perturbations up

IGF1-MCF7
HBEGF-BT20
EPG-MCF7
MCSF-MCF7
IGF-MCF7

LINCS L1000 Ligand Perturbations down

EPR-SKBR3
EGF-SKBR3
MCSF-MCF7
PDGFBB-SKBR3
HBEGF-SKBR3

ARCHS4 IDG Coexp

PRKRA_IDG_kinase_ARCHS4_coexpression
DGKH_IDG_kinase_ARCHS4_coexpression
TMEM38B_IDG_ionchannel_ARCHS4_coexpn
CDK17_IDG_kinase_ARCHS4_coexpression
GLRA3_IDG_ionchannel_ARCHS4_coexpressi

DrugMatrix

Carbon_Tetrachloride-400_mg/kg_in_Corn_C
44'-Methylenedianiline-81_mg/kg_in_Corn_C
Allopurinol-175_mg/kg_in_Corn_Oil-Rat-Kidn
Econazole-334_mg/kg_in_Corn_Oil-Rat-Liver
Cytarabine-407_mg/kg_in_Saline-Rat-Liver-Si

Old CMAP up

phenanthridinone-1115
haloperidol-2039
hexamethonium bromide-1982
diphemidol mesilsulfate-1912
BC19220-706

Old CMAP down

bupropion-1564
helveticoside-3945
strophanthidin-2525
flunixin-2552
fenoldipine-3264

GeneSigDB

17676974-Table55
17962622-Table2
17975224-Table56
16909099-SupTable2
18440302-SupTable1

OMIM Disease

glycogen_storage_disease
leigh_syndrome
cardiomyopathy_hypertrophic
neuropathy
blood

OMIM Expanded

glycogen_storage_disease
leigh_syndrome
polydactyly
cardiomyopathy_hypertrophic
complex_I

VirusMINT

Human papillomavirus type 8
Epstein-Barr virus (strain 895-8)
Bovine papillomavirus type 1
Human papillomavirus type 18
Human papillomavirus type 11

MSigDB Computational

MODULE_78
GCM_SIRT2
MODULE_43
MODULE_21
MODULE_237

MSigDB Oncogenic Signatures

AKT_UP_MTOR_DN.V1_DN
MTOR_UP.N4.V1_DN
ESC_V6.5_UP.LATE.V1_UP
MTOR_UP.N4.V1_UP
GCNP_SHH_UP_EARLY.V1_UP

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](#)

» Introduction

WebGestalt (WEB-based Gene SeT Analysis Toolkit) 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

» Basic Parameters

Select Organism of Interest [?]

Select Method of Interest [?]

Select Functional Database [?]

Gene List

Select Gene ID Type [?]

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

Reference Gene List

Select Reference Set for Enrichment Analysis [?]

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type [?]

» Advanced parameters

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 browsers. 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

Select Reference Set for Enrichment Analysis 

genome_protein-coding 

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 Entrez IDs or not mapped

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

WEBGESTAL

Summary

User ID Mapping Table

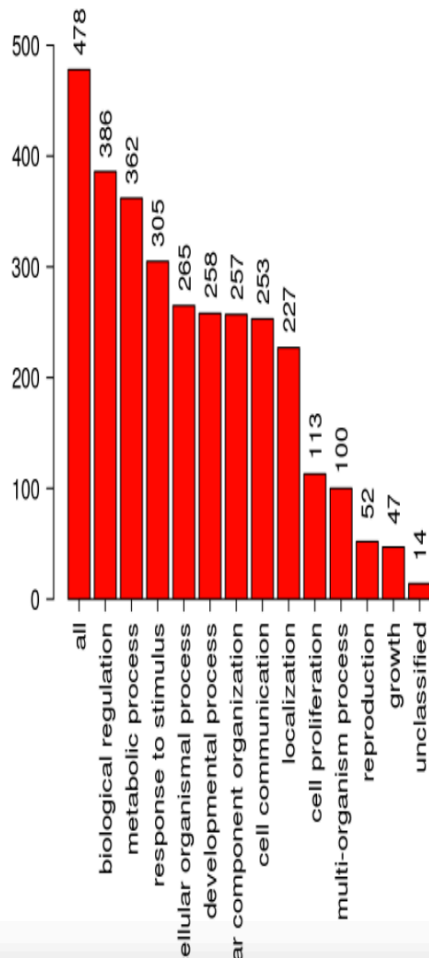
GOSlim Summary

Enrichment Results

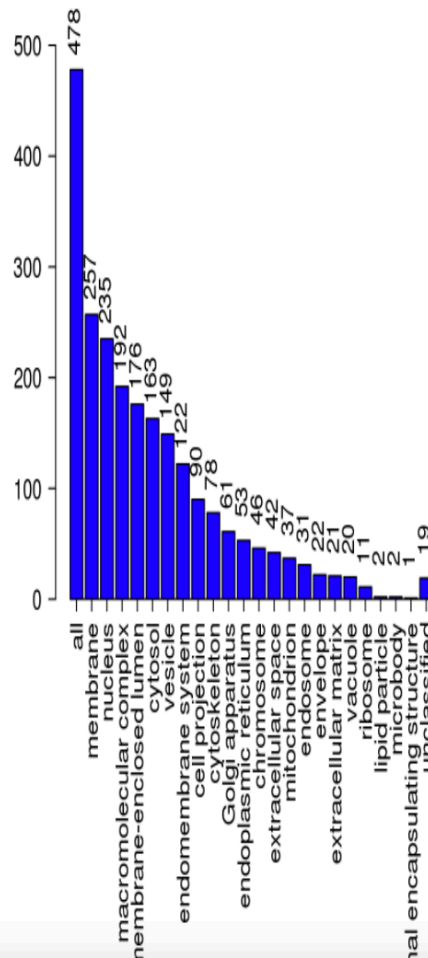
GOSlim summary for the user list genes

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

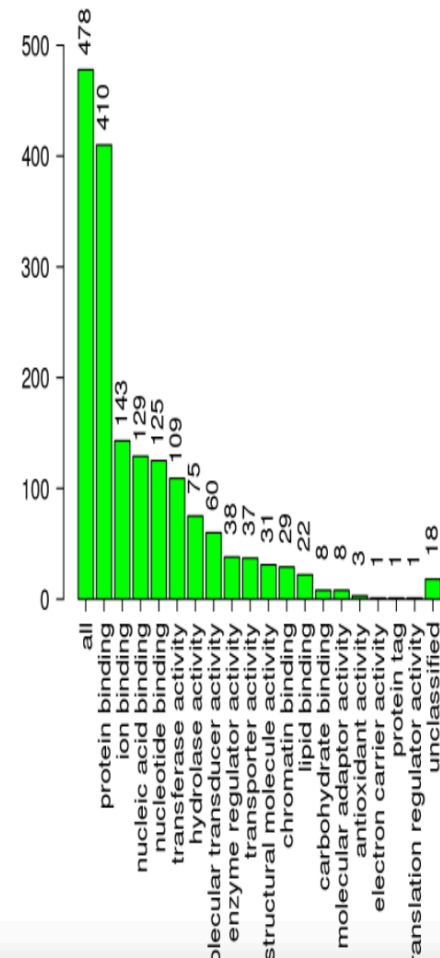
Bar chart of Biological Process categories



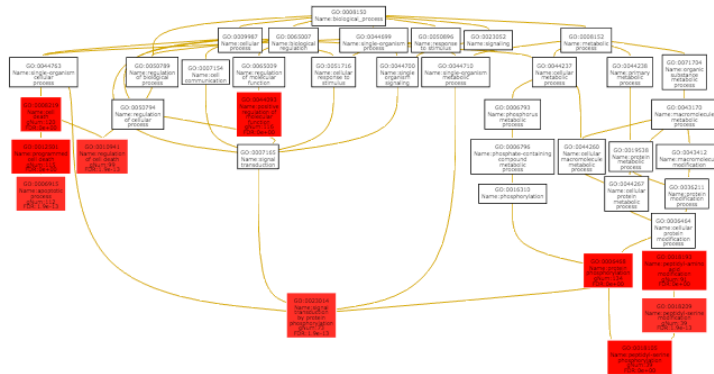
Bar chart of Cellular Component categories



Bar chart of Molecular Function categories



Enrichment Results



Detailed information of the enriched categories

The statistics **1** 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, endonuclease	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	128

KEGG

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



KEGG [Help](#)

[» Japanese](#)

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

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (See [Release notes](#) for new and updated features).

Main entry point to the KEGG web service

KEGG2 [KEGG Table of Contents](#) [Update notes](#)

Data-oriented entry points

KEGG PATHWAY [KEGG pathway maps](#) [[Pathway list](#)]
KEGG BRITE [BRITE functional hierarchies](#) [[Brite list](#)]
KEGG MODULE [KEGG modules](#) [[Module list](#)]
KEGG DISEASE [Human diseases](#) [[Cancer](#) | [Infectious disease](#)]
KEGG DRUG [Drugs](#) [[ATC drug classification](#)]
KEGG ORTHOLOGY [Ortholog groups](#) [[KO system](#)]
KEGG GENOME [Genomes](#) [[KEGG organisms](#)]
KEGG GENES [Genes and proteins](#) [Release history](#)
KEGG COMPOUND [Small molecules](#) [[Compound classification](#)]
KEGG REACTION [Biochemical reactions](#) [[Reaction modules](#)]

Entry point for wider society

KEGG MEDICUS [Health-related information resource](#)

Organism-specific entry points

KEGG Organisms Enter org code(s) [hsa](#) [hsa eco](#)

Analysis tools

KEGG Mapper [KEGG PATHWAY/BRITE/MODULE mapping tools](#)
KEGG Atlas [Navigation tool to explore KEGG global maps](#)
KAAS [KEGG automatic annotation server](#)
BLAST/FASTA [Sequence similarity search](#)
SIMCOMP [Chemical structure similarity search](#)
PathPred [Biodegradation/biosynthesis pathway prediction](#)

KEGG

KEGG2 PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET

Search 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 <i>New!</i> 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 interpretation 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

01100 Metabolic pathways

- 01110 Biosynthesis of secondary metabolites
- 01120 Microbial metabolism in diverse enviro
- 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, pha
- ☐ 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

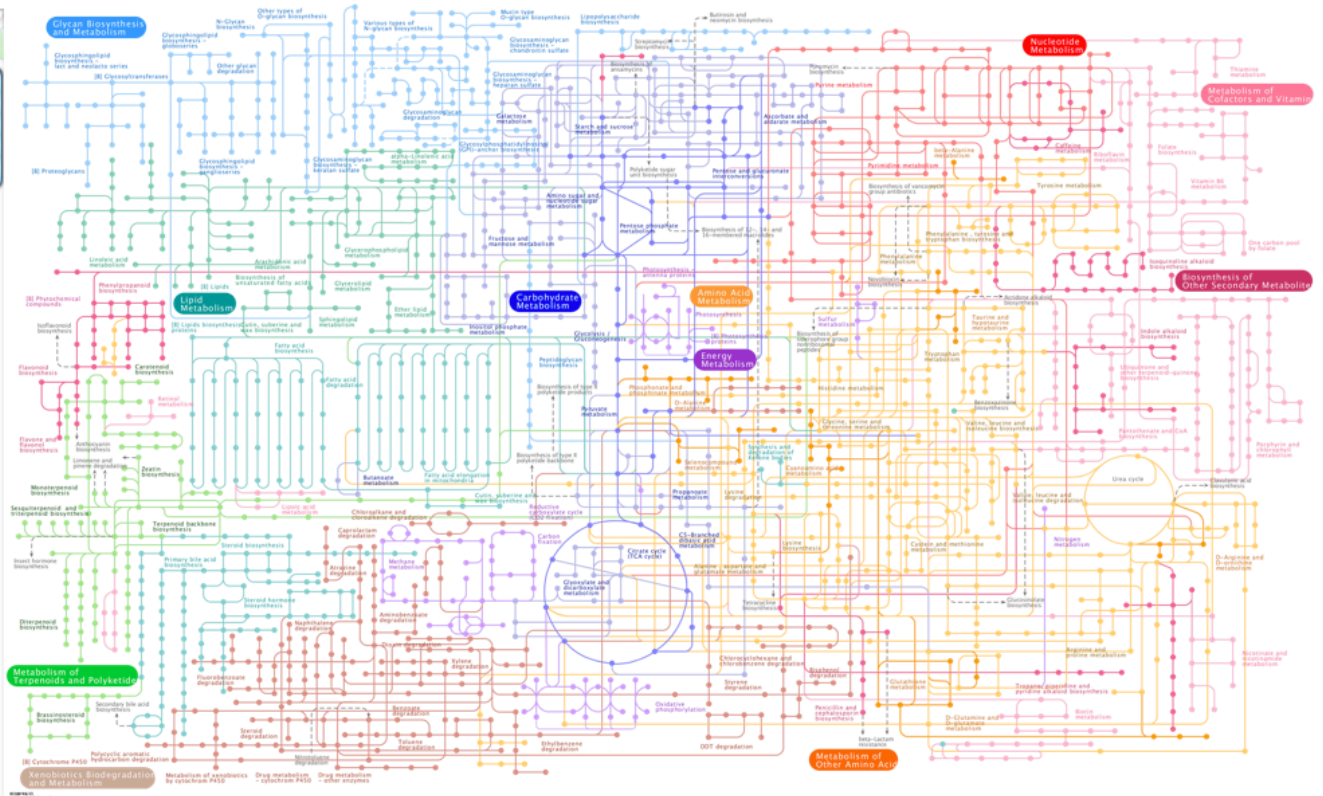
- ☐ M00176 Assimilatory sulfate reduction, s

Carbohydrate and lipid metabolism

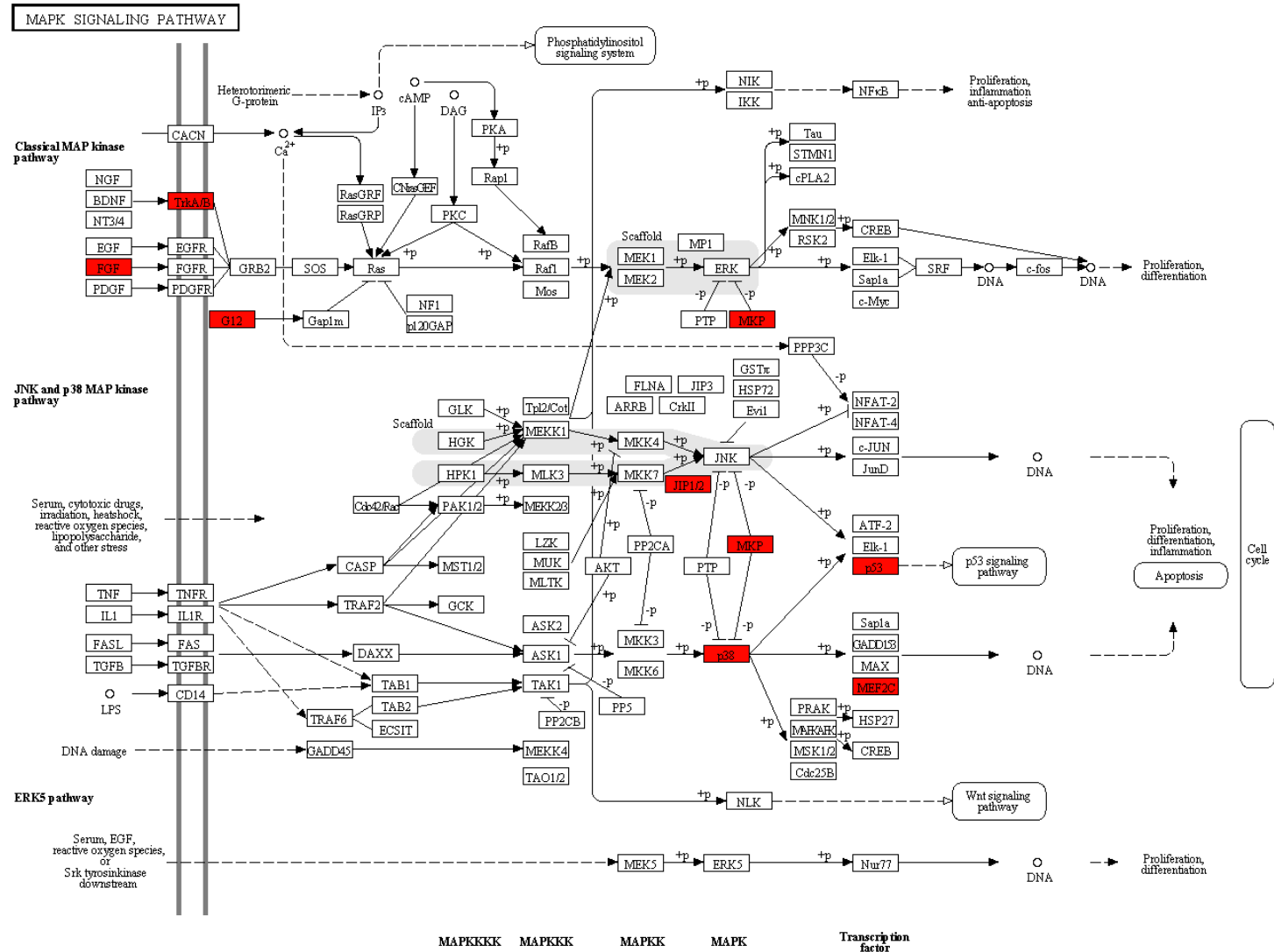
Central carbohydrate metabolism

- ☐ M00001 Glycolysis (Embden-Meyerhof p
- ☐ M00002 Glycolysis, core module involv
- ☐ 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, no

KEGG



KEGG



Other Pathway Databases



The Reactome website features a dark blue header with the Reactome logo and a navigation bar. Below the header, there are several large buttons for 'Browse Pathways', 'Analyze Data', 'Reactome FI Network', 'User Guide', 'Data Download', and 'Contact Us'. A 'Reactome News' section on the right contains a welcome message and a tweet from Ewan Birney. The footer lists supporting institutions like OICR, NYU Langone, CSH, Cold Spring Harbor Laboratory, and EMBL-EBI.

REACTOME
A CURATED PATHWAY DATABASE

About Content Documentation Tools Community Download Contact

Browse Pathways Analyze Data Reactome FI Network User Guide Data Download Contact Us

Reactome News

Welcome to the New Reactome Website

reactome @reactome 1 Oct
Reactome receives a ORCID Adoption & Integration Program Award: Reactome is one of nine project partners who w... [bit.ly/18MuGcF](#)
Expand

Ewan Birney @ewanbirney 1 Oct
Absolutely loving the new @reactome web site [reactome.org](#). Molecular pathways with polish and accuracy! Hats off to the team!
Retweeted by reactome
Expand

reactome @reactome 1 Oct
One of the new pathways from V46 release, Signaling by the TGF-beta Receptor Complex in Cancer, [goo.gl/YN4cKp](#), #userreactome

Tweet to @reactome

The development of Reactome is supported by a grant from the US National Institutes of Health (P41 HG003751), Ontario Research Fund, and the European Molecular Biology Laboratory.

<http://www.reactome.org/>



The Pathway Commons website has a light blue header with the Pathway Commons logo and a search bar. Below the header, there are navigation links for Home, Data Sources, Download, FAQ, Web Service, and About. A 'Search Pathway Commons' section includes a search bar and a 'Find Pathways' button. A 'What's New' section lists recent updates, including the May 28, 2013 update and the October 27, 2011 update. A 'Using Pathway Commons' section provides information for biologists, computational biologists, and software developers. A 'Current Data Sources' section lists various data sources like BioGRID, HumanCyc, IntAct, MINT, and Reactome. A 'Pathway Commons Quick Stats' section displays statistics on the number of pathways, interactions, and entities. A 'Citing Pathway Commons' section provides information on how to cite the project and the software.

Pathway Commons

Search and visualize public biological pathway information. Single point of access. [more...]

Home Data Sources Download FAQ Web Service About

Send us your feedback. Sign up for Pathway Commons announcements. RSS Feed

Search Pathway Commons:

Find Pathways Find Molecules

For example, if you enter: [BRCA1](#), you will get back the list of pathways containing the keyword "BRCA1", and the list of pathways that contain the BRCA1 gene.

Current filter settings: All Organisms, All Data Sources. [Set filters](#).

Using Pathway Commons:

Biologists: Browse and search pathways across multiple valuable public pathway databases.

Computational biologists: Download an integrated set of pathways in BioPAX format for global analysis.

Software developers: Build software on top of Pathway Commons using our [web service API](#). Download and install the [cPath software](#) to create a local mirror.

Current Data Sources:

Pathway Commons currently contains the following data sources ([batch download](#)):

BioGRID Metabolic State-Transition Cancer Center Human Protein Reference Database

HUMANCYC SBGRID IntAct

MINT Reactome

Pathway Commons Quick Stats:

Number of Pathways: 1,668
Number of Interactions: 442,182
Number of Physical Entities: 86,282
Number of Organisms: 414

Integration of additional data sources is planned in the near future. For a comprehensive directory of interaction and pathway databases, please refer to [Pathguide](#).

Citing Pathway Commons:

To cite the Pathway Commons Project: Cerami et al. Pathway Commons, a web resource for biological pathway data. Nucl. Acids Res. (2010) doi:10.1093/nar/gkq1039

To cite the cPath Software: Cerami et al. cPath: open source software for collecting, storing, and querying biological pathways. BMC Bioinformatics. (2006) doi:10.1186/1471-2105-7-497

What's New:

- May 28, 2013:
 - This portal **will not be updated** in the future and we are working on a replacement Pathway Commons service for expected release in mid-2013.
 - We encourage you to test the [new service](#) and switch to it once it is released.
 - Please [send us](#) questions and feedback.
- Oct 27, 2011:
 - BioGRID data set (September 25, 2011 Version 3.1.81).
 - IntAct data set (September 29, 2011 Version 3.1.17288).
 - Nature Pathway Interaction data set (October 12, 2011).
 - Reactome data set (September 20, 2011 Version 38).
- June 24, 2011:
 - BioGRID data set (May 1, 2011 Version 3.1.76).
 - HumanCyc data set (June 8, 2011 Version 15.1).
 - Nature Pathway Interaction data set (June 14, 2011).
- April 25, 2011:
 - Reactome data set (March 15, 2011 Version 36).
 - IntAct data set (February 3, 2011 Version 138).
 - MetaCyc removed due to organism generic pathways - will be brought back when these pathways are

<http://www.pathwaycommons.org/>

STRING

Version: 10.5

LOGIN | REGISTER

 <http://string-db.org/> [Search](#) [Download](#) [Help](#) [My Data](#)

Welcome to STRING

Protein-Protein Interaction Networks

ORGANISMS	PROTEINS	INTERACTIONS
2031	9.6 mio	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

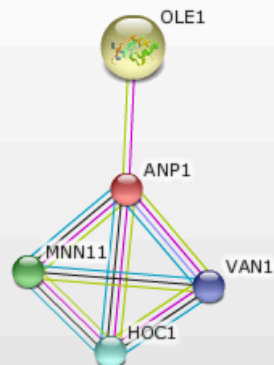
CREDITS

Funding

Datasources

Partners

Software



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*)

Predicted Functional Partners:

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

Views:



Neighborhood



Fusion



Occurrence



Coexpression



Experiments



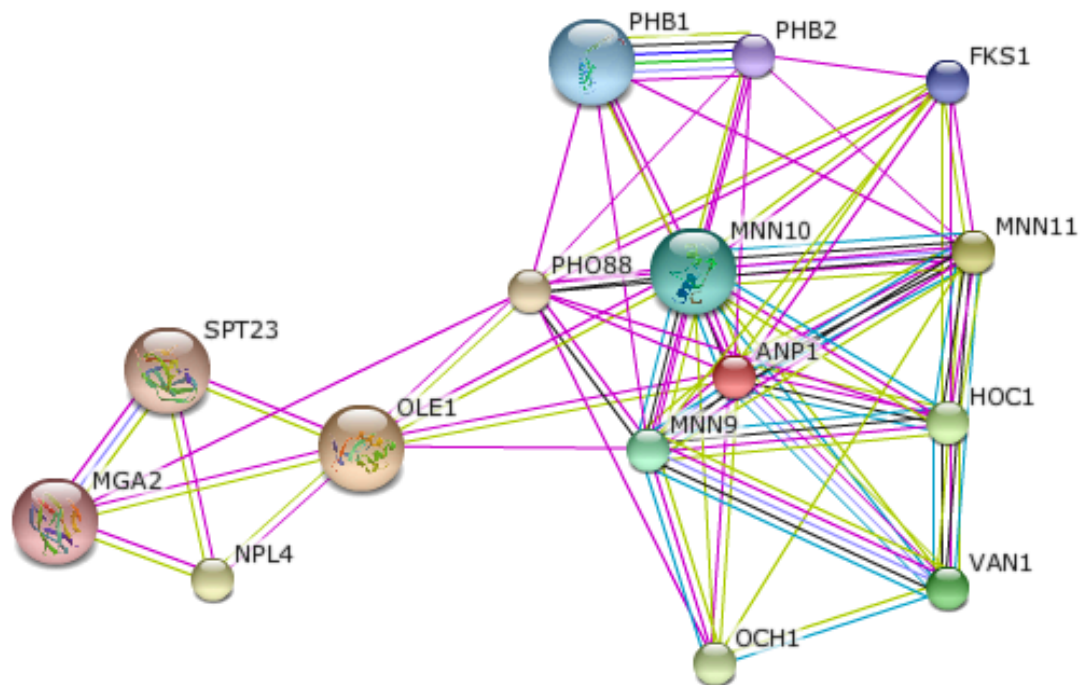
Database



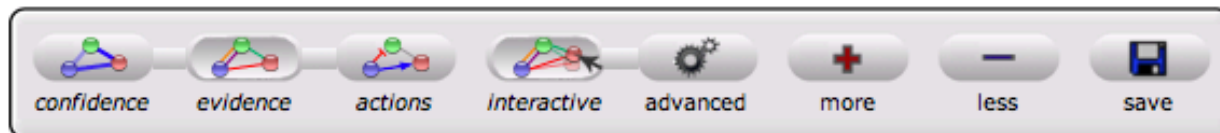
Textmining



Summary Network



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



(requires Flash player 10 or better)

Cytoscape



The screenshot shows the Cytoscape website homepage. At the top, there is a navigation bar with links: Home, Introduction, Download, Apps, Documentation, Community, Report a Bug, and Getting Help. A search bar is located in the top right corner. The main content area features a large graphic of the Cytoscape logo (a network of orange nodes and edges) and the word "Cytoscape" in a large, bold, black and orange font. To the right of the logo, there is a text box titled "Network Data Integration, Analysis, and Visualization in a Box" which describes Cytoscape as an open source software platform for visualizing complex networks and integrating them with any type of attribute data. Below this text box are four buttons: "Download Cytoscape" (orange), "Welcome Letter" (green), "Release Notes" (green), and "Sample Visualizations" (blue). At the bottom of the page, there is a section titled "Learning Cytoscape" with the URL "http://www.cytoscape.org/" and four buttons: "Introduction to Cytoscape" (orange), "Tutorial for 3.x" (orange), "Tutorial for 2.x" (green), and "Developers Tutorial" (grey).

Cytoscape

Search... Go

Home Introduction Download Apps Documentation Community Report a Bug Getting Help

 **Cytoscape**

i

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

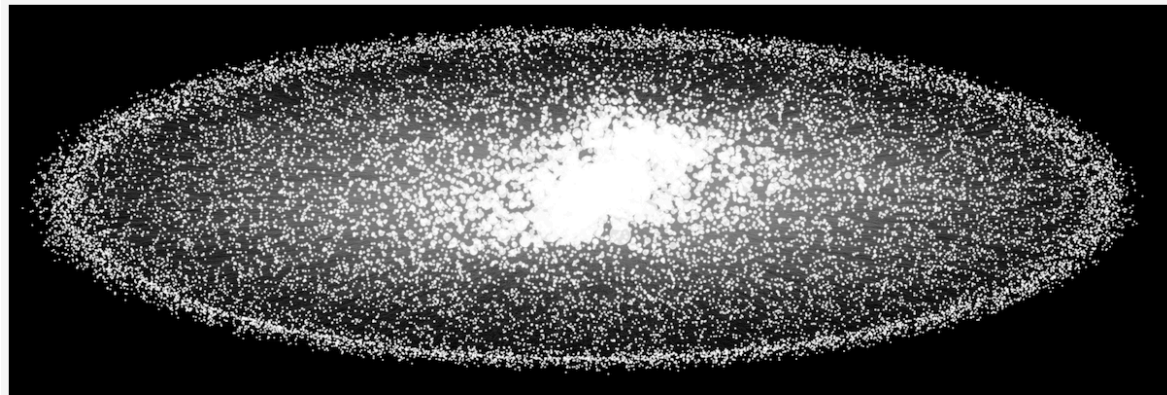
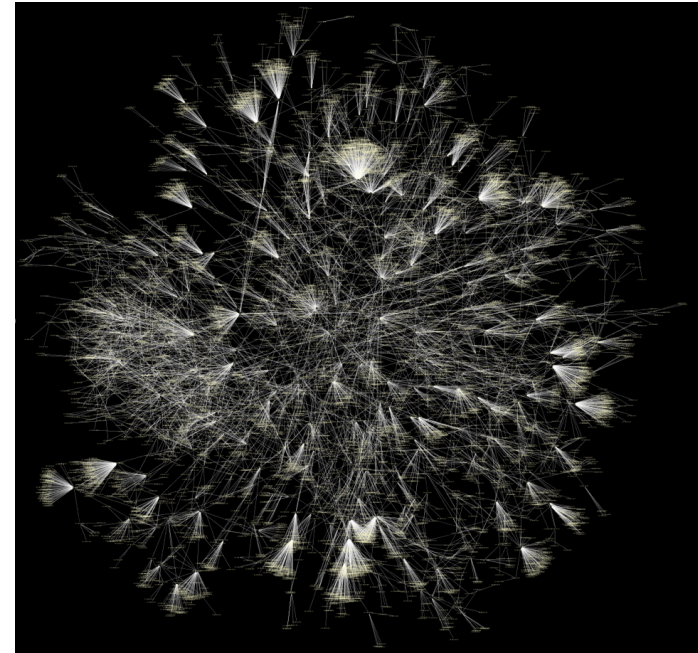
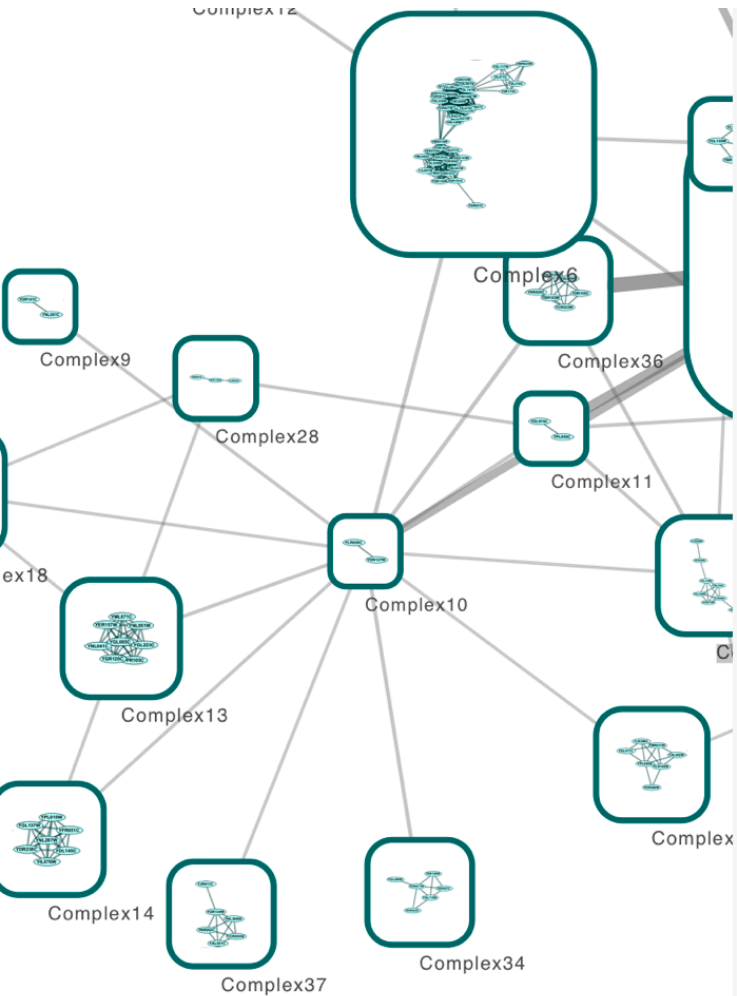
Release Notes

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

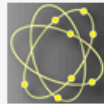
PSICQUIC Web Service Client for importing interactions from public

BiNGO

BiNGO

Calculates overrepresented GO terms in the network and display

3.0+



ClueGO

Creates and visualizes a functionally grouped network of

3.0+

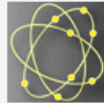


DynNetwork

Visualize dynamic networks in Cytoscape 3.0

3.0+

Top Voted Apps



ClueGO

Creates and visualizes a functionally grouped network of

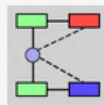
3.0+

GeneMANIA

GeneMANIA

Imports interaction networks from public databases from a list of

3.0+



CytoKegg

Identify Kegg pathways associated to specific

3.0+

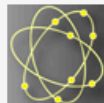


PathExplorer

Finds paths, filters them based on node and edge attributes and

3.0+

Top Downloaded Apps



ClueGO

Creates and visualizes a functionally grouped network of

3.0+

GeneMANIA

GeneMANIA

Imports interaction networks from public databases from a list of

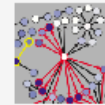
3.0+



MCODE

Clusters a given network based on topology to find densely

3.0+

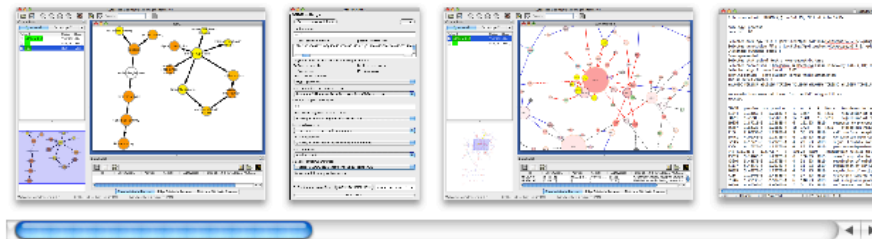


jActiveModules

Finds clusters where member nodes show significant changes

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](#))

BEReX: Biomedical Entity-Relationship eXplorerMinji Jeon¹, Sunwon Lee², Kyubum Lee², Aik-Choon Tan³ and Jaewoo Kang^{1,2,*}¹Interdisciplinary Graduate Program in Bioinformatics, ²Department of Computer Science and Engineering, Korea University, Seoul, Korea and ³Department of Medicine/Medical Oncology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Associate Editor: Mario Albrecht

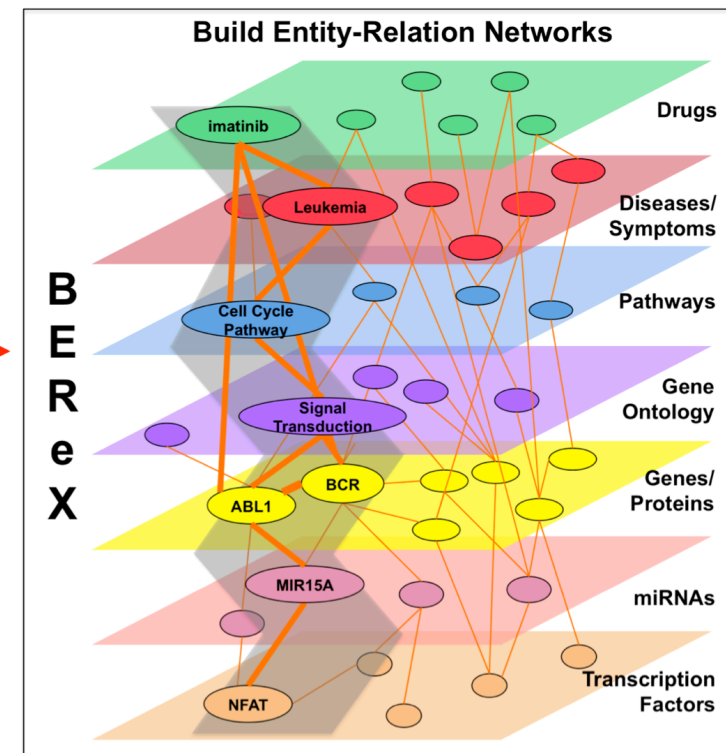
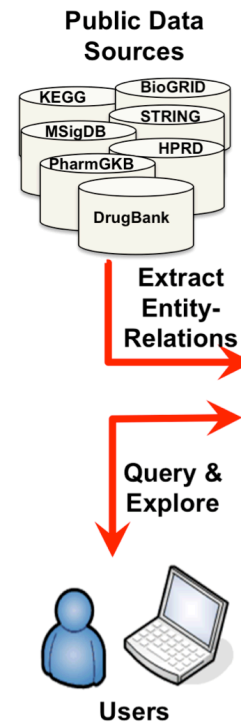
**ABSTRACT**

Summary: Biomedical Entity-Relationship eXplorer (BEReX) is a new biomedical knowledge integration, search and exploration tool. BEReX integrates eight popular databases (STRING, DrugBank, KEGG, PharmGKB, 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.

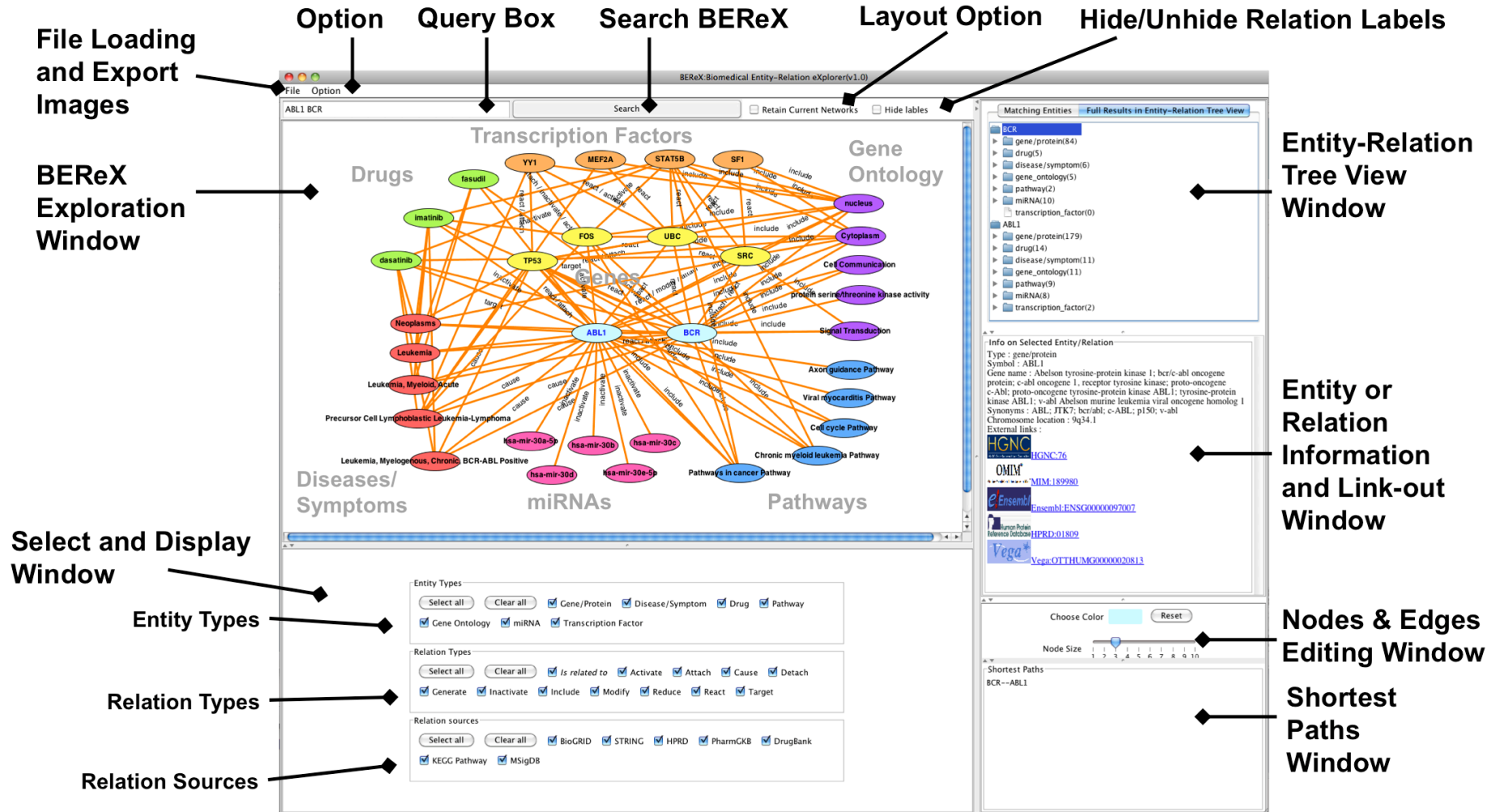
Availability and implementation: BEReX and a detailed user guide are available for download at our project Web site (<http://infos.korea.ac.kr/berex>).

Contact: kangj@korea.ac.kr

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



BEReX: Biomedical Entity-Relationship eXplorer



(Collaboration with Dr. Jaewoo Kang, Korea University)

iBEReX (<http://iberex.korea.ac.kr/>)

Query Analysis Help

BCR
ABL1
imatinib

submit

Example1 Example2

Default color scheme

- Gene (yellow) ○ Disease (white) ● Pathway (blue)
- Drug (sky blue) ● GO term (purple) ● miRNA (gray)
- Transcription factor (orange)
- Overexpressed gene (red) ● Underexpressed gene (green)

Query by File Upload

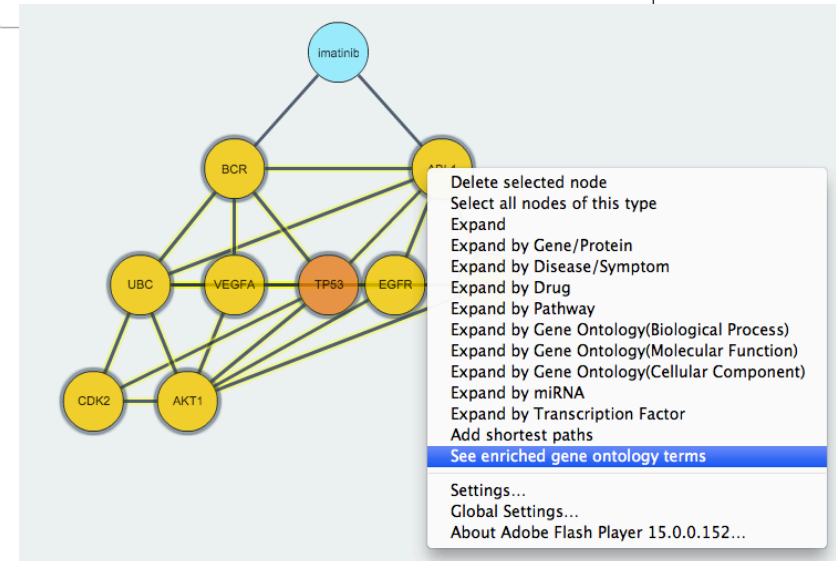
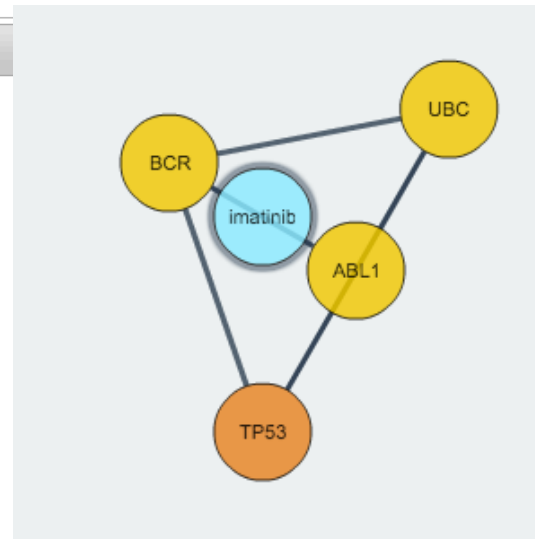
Choose File No file chosen

submit

Example

querylist.txt

submit



Entity Relationship Tree Entity Relationship Type Node/Edge Info Graph Edit

Enriched Gene Ontology Terms

Result of BCR, TP53, AKT1, CDK2, UBC, APP, VEGFA, EGFR, ABL1

Biological process
transmembrane transport, multicellular organismal development, negative regulation of transcription, DNA-dependent, intracellular signal transduction, negative regulation of cell proliferation, transcription, DNA-dependent, carbohydrate metabolic process, G-protein coupled receptor signaling pathway, extracellular matrix organization, inflammatory response, response to drug, regulation of transcription from RNA polymerase II promoter, protein ubiquitination

Molecular Function
zinc ion binding, sequence-specific DNA binding transcription factor activity, protein homodimerization activity, metal ion binding, receptor binding, protein kinase activity

Chemical Compound
extracellular vesicular exosome, integral to membrane, integral to plasma membrane, endoplasmic reticulum membrane, endoplasmic reticulum, Golgi apparatus, Golgi membrane, extracellular region, protein complex, centrosome

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.