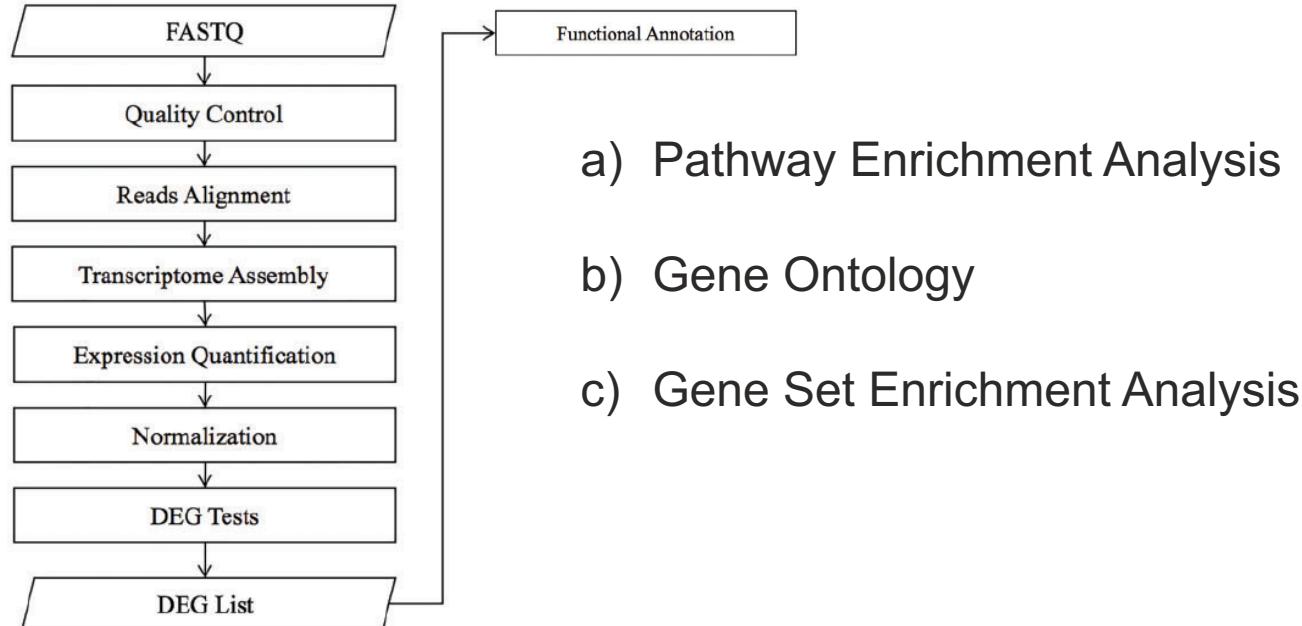


Pathway and Gene Set Enrichment Analysis

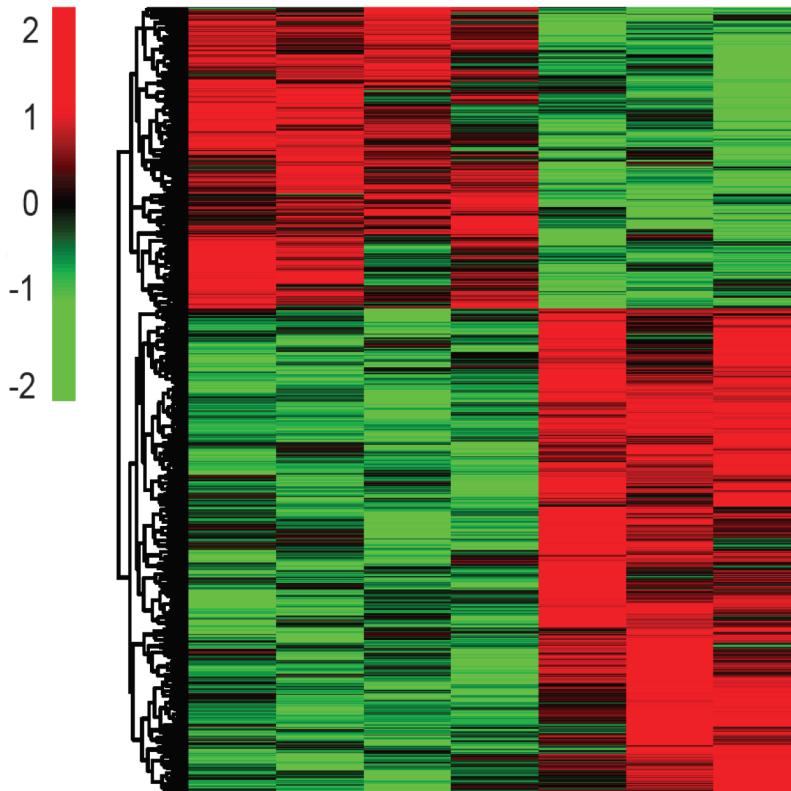
Parveen Kumar



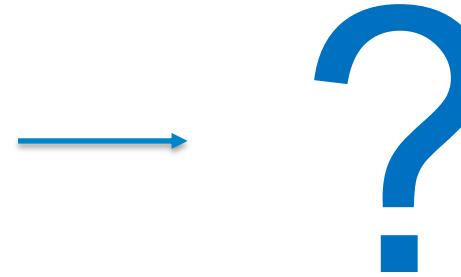
Gene Set Enrichment Analysis



Pathway Enrichment Analysis



Heatmap of Differential
Genes (DGEs)



What pathways are enriched
in these DGEs?

Why do we need pathway information?

- Understand the biological phenomena involved
- To understand the gene network (Genes do not work alone, but in an intricate network of interactions)
- In-depth and contextualized findings to help understand the mechanisms of disease in question
- Prediction of drug targets
- Understand how to intervene therapeutically in disease processes
- Conduct targeted literature searches

Online Tools

- DAVID
 - <https://david.ncifcrf.gov/home.jsp>

The screenshot shows the DAVID Bioinformatics Resources 6.8 homepage. At the top, it says "DAVID Bioinformatics Resources 6.8" and "Laboratory of Human Retrovirology and Immunobiology (LRIU)". Below that is a banner with the text "*** If you are looking for DAVID 6.7, please visit our development site. ***" and "Recommend: A paper published in *Nature Protocols* describes step-by-step procedure to use DAVID". The main content area has sections for "Functional Annotation", "Gene Functional Classification", "Gene Name Batch Viewer", and "Gene ID Cross-viewer". On the right side, there's a sidebar titled "What's Important in DAVID" with links to "DAVID", "DAVID 6.8", "Novel Classification Algorithms", "DAVID High Throughput", and "Enhanced gene background". Another sidebar titled "Statistics of DAVID" shows a bar chart of "DAVID Citations (2003-2017)".

- Reactome
 - <https://reactome.org/>
- Enrichr
 - <http://amp.pharm.mssm.edu/Enrichr/>

The screenshot shows the Reactome homepage with the title "reactome" and the subtitle "Find Reactions, Proteins and Pathways". It features a search bar with the placeholder "e.g. C59621, HIF1, signaling by EGFR, glucose". Below the search bar are four main icons: "Pathway Browser" (visualize and interact with Reactome biological pathways), "Analyze Data" (discover enriched functional-related gene groups, cluster redundant annotation terms, visualize genes on BioCarta & KEGG pathway maps), "ReactomeFIViz" (design to find pathways and network patterns related to cancer and other types of diseases), and "Documentation" (information to browse the database and use its principal tools for data analysis).

The screenshot shows the Enrichr homepage. At the top, it says "Enrichr" and "9,558,842 Bets Analyzed, 238,875 terms, 129 Libraries". It has tabs for "Analyze", "What's New?", "Libraries", "Find a Gene", "About", and "Help". The main section is titled "Input data" with instructions: "Choose an input file to be uploaded. 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 exclude the gene from the enrichment analysis and the weight of 1.0 is the maximum." There is a "Try an example BED file" link and a "Choose File" button. Below that is a large input field for "0 gene(s) entered" and a "Submit" button.

DAVID Tool for Functional Annotation

DAVID Bioinformatics Resources 6.8
Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

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

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

Recommending: A paper published in *Nature Protocols* describes step-by-step procedure to use DAVID!

Welcome to DAVID 6.8

2003 - 2018

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 Citations (2003-2017)

Year	Citations
03	~10
04	~20
05	~40
06	~60
07	~100
08	~150
09	~250
10	~400
11	~600
12	~800
13	~1200
14	~1800
15	~2500
16	~3200
17	~3800

- ≥ 33,000 Citations
- Average Daily Usage: ~2,700 gene lists/sublists from >900 unique researchers.
- Average Annual Usage: ~1,000,000 gene lists/sublists from >100 countries

Screen Shot 1 Screen Shot 2 Screen Shot 3

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

DAVID: Submitting Gene List

- Paste Gene list
- Choose gene list
E.g. Official Gene Symbol
- Select List Type
e.g. Gene List

The screenshot shows the DAVID Bioinformatics Resources 6.8 interface. A red oval highlights the 'Upload Gene List' section. Three blue arrows point from the user instructions on the left to specific fields in this section: one arrow points to the 'A: Paste a list' text input field; another points to the 'B: Choose From a File' section; and a third points to the 'Step 3: List Type' radio button options.

Functional Annotation Tool
DAVID Bioinformatics Resources 6.8, 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 ***
*** If you are looking for DAVID 6.7, please visit our [development site](#). ***

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.

DAVID: Select Species and View results

DAVID BIOINFORMATICS DATABASE Functional Annotation Tool DAVID Bioinformatics Resources 6.8, 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 ***
*** 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 -
Mus musculus(65)
Rattus norvegicus(65)
Canis lupus familiaris(64)

Select Species

List Manager Help

List_1

Select List to:
Use Rename
Remove Combine
Show Gene List

Annotation Summary Results

Help and Tool Manual

Current Gene List: List_1
Current Background: Mus musculus

65 DAVID IDs Check Defaults

- Functional_Categories** (3 selected)
- Gene_Ontology** (3 selected)
- General_Annotations** (0 selected)
- Literature** (0 selected)
- Main_Accessions** (0 selected)
- Pathways** (2 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

Select Species
e.g. Mouse

DAVID: Functional Annotation Chart



DAVID Bioinformatics Resources 6.8
Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

*** Welcome to DAVID 6.8 ***

*** If you are looking for DAVID 6.7, please visit our development site. ***

Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List_1

Current Background: Mus musculus

65 DAVID IDs

Options

[Rerun Using Options](#) [Create Sublist](#)

240 chart records

[Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	KEGG_PATHWAY	GABAergic synapse	RT		12	18.5	1.4E-14	1.3E-12
<input type="checkbox"/>	KEGG_PATHWAY	Morphine addiction	RT		10	15.4	5.8E-11	2.8E-9
<input type="checkbox"/>	GOTERM_CC_DIRECT	heterotrimeric G-protein complex	RT		7	10.8	2.0E-9	2.5E-7
<input type="checkbox"/>	UP_KEYWORDS	Synapse	RT		12	18.5	4.4E-9	6.0E-7
<input type="checkbox"/>	GOTERM_CC_DIRECT	synapse	RT		14	21.5	4.8E-9	3.1E-7
<input type="checkbox"/>	KEGG_PATHWAY	Glutamatergic synapse	RT		9	13.8	1.1E-8	3.6E-7
<input type="checkbox"/>	UP_KEYWORDS	Lipoprotein	RT		15	23.1	2.8E-8	1.9E-6
<input type="checkbox"/>	KEGG_PATHWAY	Dopaminergic synapse	RT		9	13.8	3.8E-8	8.9E-7
<input type="checkbox"/>	GOTERM_CC_DIRECT	membrane	RT		44	67.7	6.4E-8	2.7E-6
<input type="checkbox"/>	KEGG_PATHWAY	Circadian entrainment	RT		8	12.3	8.8E-8	1.7E-6
<input type="checkbox"/>	KEGG_PATHWAY	Retrograde endocannabinoid signaling	RT		8	12.3	1.2E-7	2.0E-6
<input type="checkbox"/>	INTERPRO	G-protein gamma-like domain	RT		5	7.7	2.4E-7	4.2E-5
<input type="checkbox"/>	UP_KEYWORDS	Prenylation	RT		8	12.3	3.0E-7	1.3E-5
<input type="checkbox"/>	SMART	GGL	RT		5	7.7	3.2E-7	1.5E-5
<input type="checkbox"/>	UP_KEYWORDS	Cell junction	RT		12	18.5	2.2E-6	7.3E-5
<input type="checkbox"/>	KEGG_PATHWAY	Cholinergic synapse	RT		7	10.8	4.7E-6	6.4E-5
<input type="checkbox"/>	KEGG_PATHWAY	Serotonergic synapse	RT		7	10.8	1.2E-5	1.4E-4

DAVID: Enriched Genes

 National Institute of Allergy and Infectious Disease
The Database for Annotation, Visualization and Integrated Discovery

Gene Report

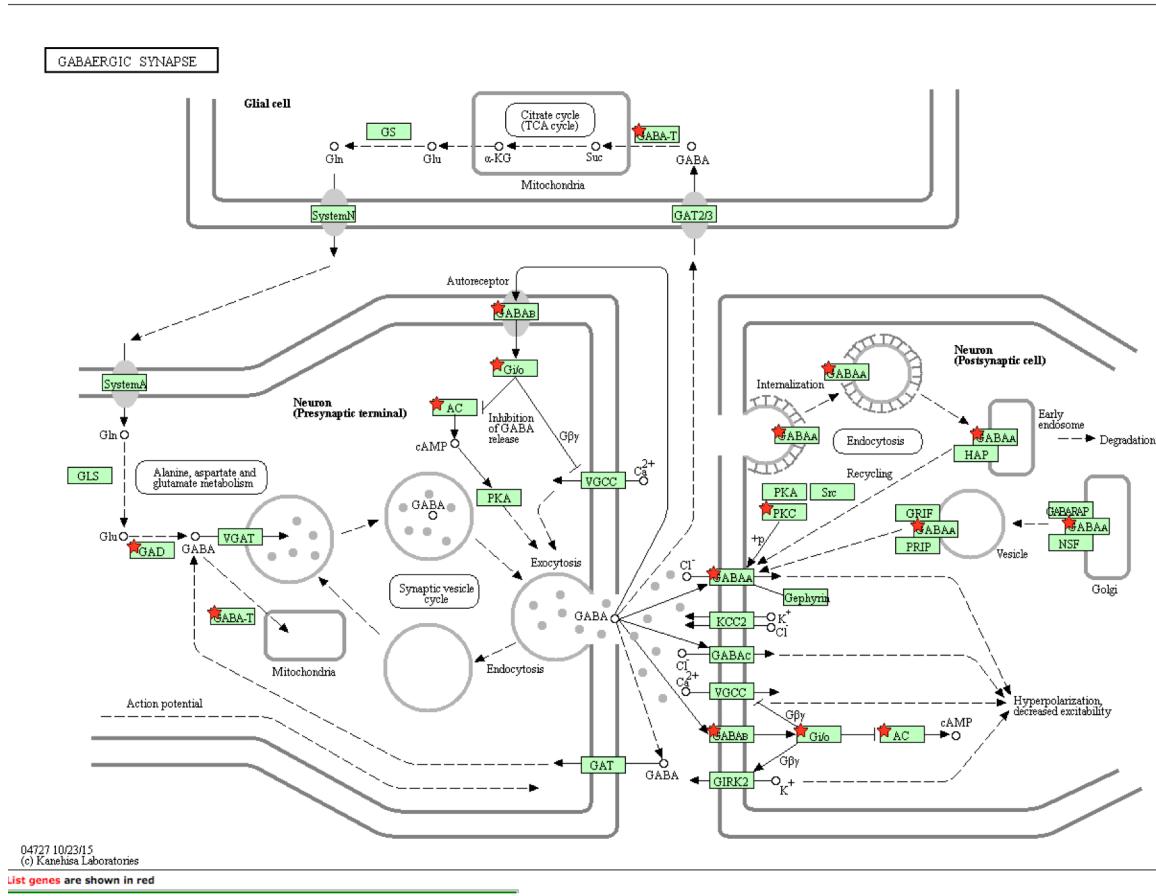
[Help and Manual](#)

Current Gene List: List_1
Current Background: *Mus musculus*
65 DAVID IDs
12 record(s)

[!\[\]\(6647c4124a390144efd78e5ada46601c_img.jpg\) Download File](#)

OFFICIAL_GENE_SYMBOL	GENE NAME	Related Genes	Species
Abat	4-aminobutyrate aminotransferase(Abat)	RG	Mus musculus
Adcy1	adenylate cyclase 1(Adcy1)	RG	Mus musculus
Adcy5	adenylate cyclase 5(Adcy5)	RG	Mus musculus
Gabra1	gamma-aminobutyric acid (GABA) A receptor, subunit alpha 1(Gabra1)	RG	Mus musculus
Gabbr2	gamma-aminobutyric acid (GABA) B receptor, 2(Gabbr2)	RG	Mus musculus
Gad1	glutamate decarboxylase 1(Gad1)	RG	Mus musculus
Gad2	glutamic acid decarboxylase 2(Gad2)	RG	Mus musculus
Gng13	guanine nucleotide binding protein (G protein), gamma 13(Gng13)	RG	Mus musculus
Gng2	guanine nucleotide binding protein (G protein), gamma 2(Gng2)	RG	Mus musculus
Gng3	guanine nucleotide binding protein (G protein), gamma 3(Gng3)	RG	Mus musculus
Gng7	guanine nucleotide binding protein (G protein), gamma 7(Gng7)	RG	Mus musculus
Prkcg	protein kinase C, gamma(Prkcg)	RG	Mus musculus

DAVID: Visualize genes in the enriched pathways



DAVID: Download gene lists

ID	Gene Name	Species
Abat	4-aminobutyrate aminotransferase(Abat)	Mus musculus
Adcy1	adenylate cyclase 1(Adcy1)	Mus musculus
Adcy5	adenylate cyclase 5(Adcy5)	Mus musculus
Gabral	gamma-aminobutyric acid (GABA) A receptor, subunit alpha 1(Gabral)	Mus musculus
Gabbr2	gamma-aminobutyric acid (GABA) B receptor, 2(Gabbr2)	Mus musculus
Gad1	glutamate decarboxylase 1(Gad1)	Mus musculus
Gad2	glutamic acid decarboxylase 2(Gad2)	Mus musculus
Gng13	guanine nucleotide binding protein (G protein), gamma 13(Gng13)	Mus musculus
Gng2	guanine nucleotide binding protein (G protein), gamma 2(Gng2)	Mus musculus
Gng3	guanine nucleotide binding protein (G protein), gamma 3(Gng3)	Mus musculus
Gng7	guanine nucleotide binding protein (G protein), gamma 7(Gng7)	Mus musculus
Prkcg	protein kinase C, gamma(Prkcg)	Mus musculus

Reactome



About ▾ Content ▾ Docs ▾ Tools ▾ Community ▾ Download

Find Reactions, Proteins and Pathways

e.g. O95631, NTN1, signaling by EGFR, glucose

Go!



Pathway Browser

Visualize and interact with Reactome biological pathways



Analyze Data

Merges pathway identifier mapping, over-representation, and expression analysis



ReactomeFIViz

Designed to find pathways and network patterns related to cancer and other types of diseases



Documentation

Information to browse the database and use its principal tools for data analysis.

USE REACTOME GRAPH DATABASE IN YOUR PROJECT

LEARN MORE

Why Reactome

Reactome is a free, open-source, curated and peer-reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education. The current version (v64) of Reactome was released on March 26, 2018.

If you use Reactome in Asia, we suggest using our Chinese mirror site at reactome.ncpsb.org.

EMBL-EBI NYU Langone Health OICR

The development of Reactome is supported by grants from the US National Institutes of Health (P41 HG003751 and 1U54GM114833-01), Ontario Research Fund, and the European Molecular Biology Laboratory.

Tweets

reactome Retweeted @Ensembl

The interactive @reactome display highlights the #histone H3-related pathway that includes UTY (aka KDM6C). Use the search box to find our #Geneoftheweek! buff.ly/2KdFW8M

Latest News

Version 64 Released

Reactome: Submit Gene List

Analysis tools

Your data Options Analysis

Step 1: Select a file from your computer or paste your own data and click on the corresponding "Continue" button.

Select data file for analysis: No file chosen

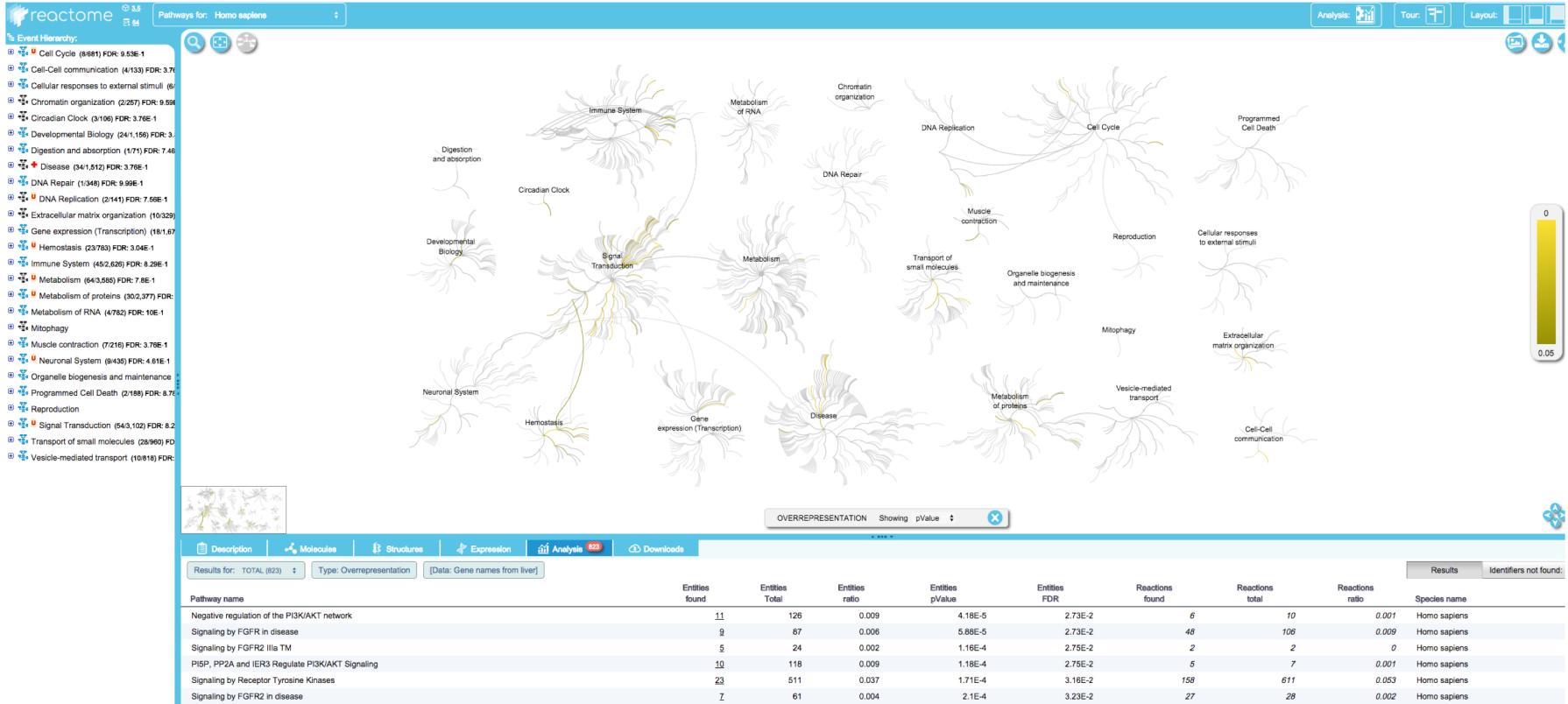
Paste your data to analyse or try example data sets:

```
#Gene names from liver
A2M
ABCB4
ABCL8
ABCB9
ABCC2
ABLM3
ACALQ2
ACACA
ACADS8
ACOT12
ACVR2C
ACVRL1
AFM
ALAS2
ALDH4A1
ANLN
ANO6
AQP7
ARHGAP26
```

Some examples:

-
-
-
-
-
-
-
-
-

Reactome: Analyze Results



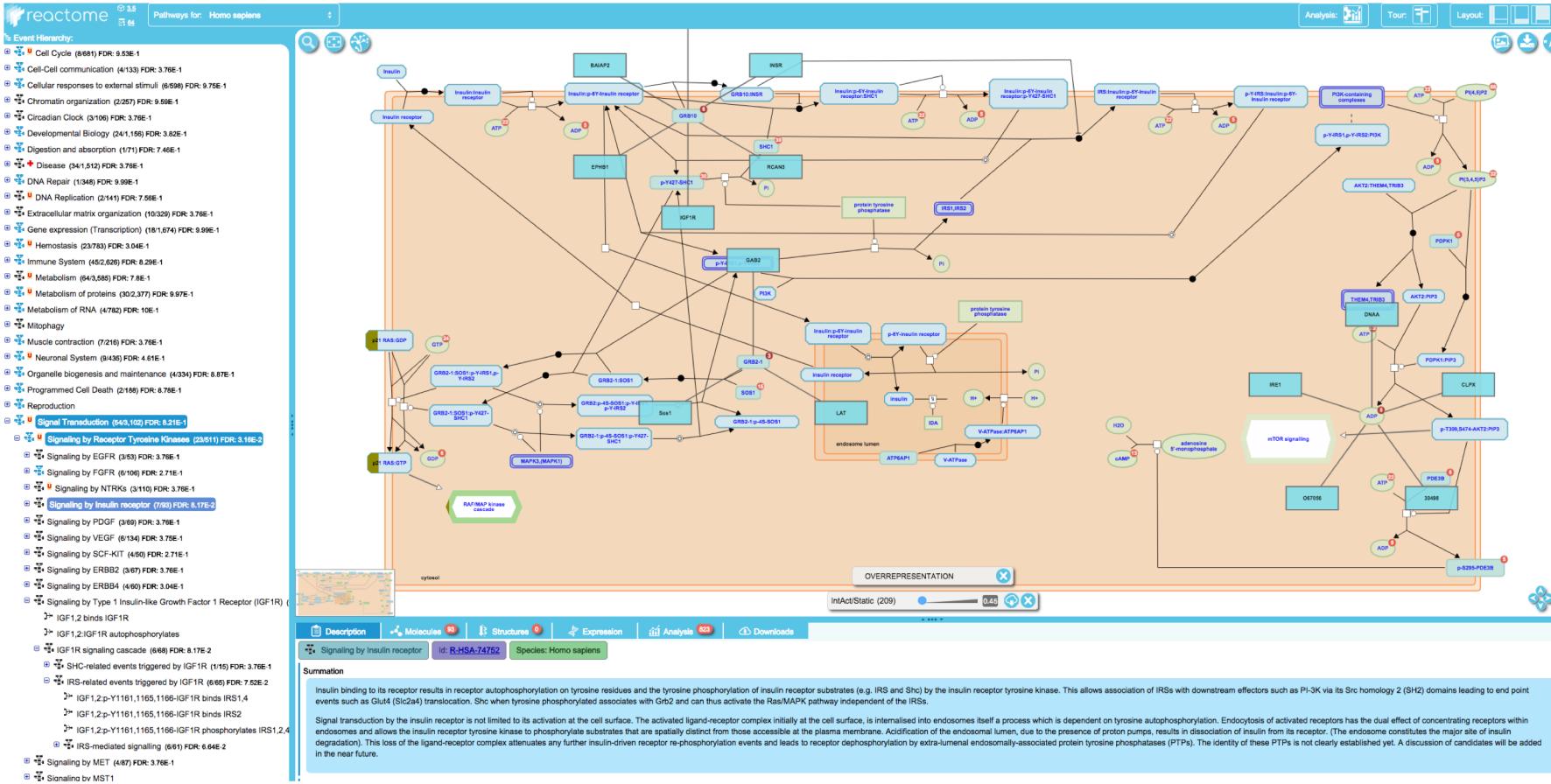
Reactome: Select specific pathway e.g. Insulin receptor

Event Hierarchy:

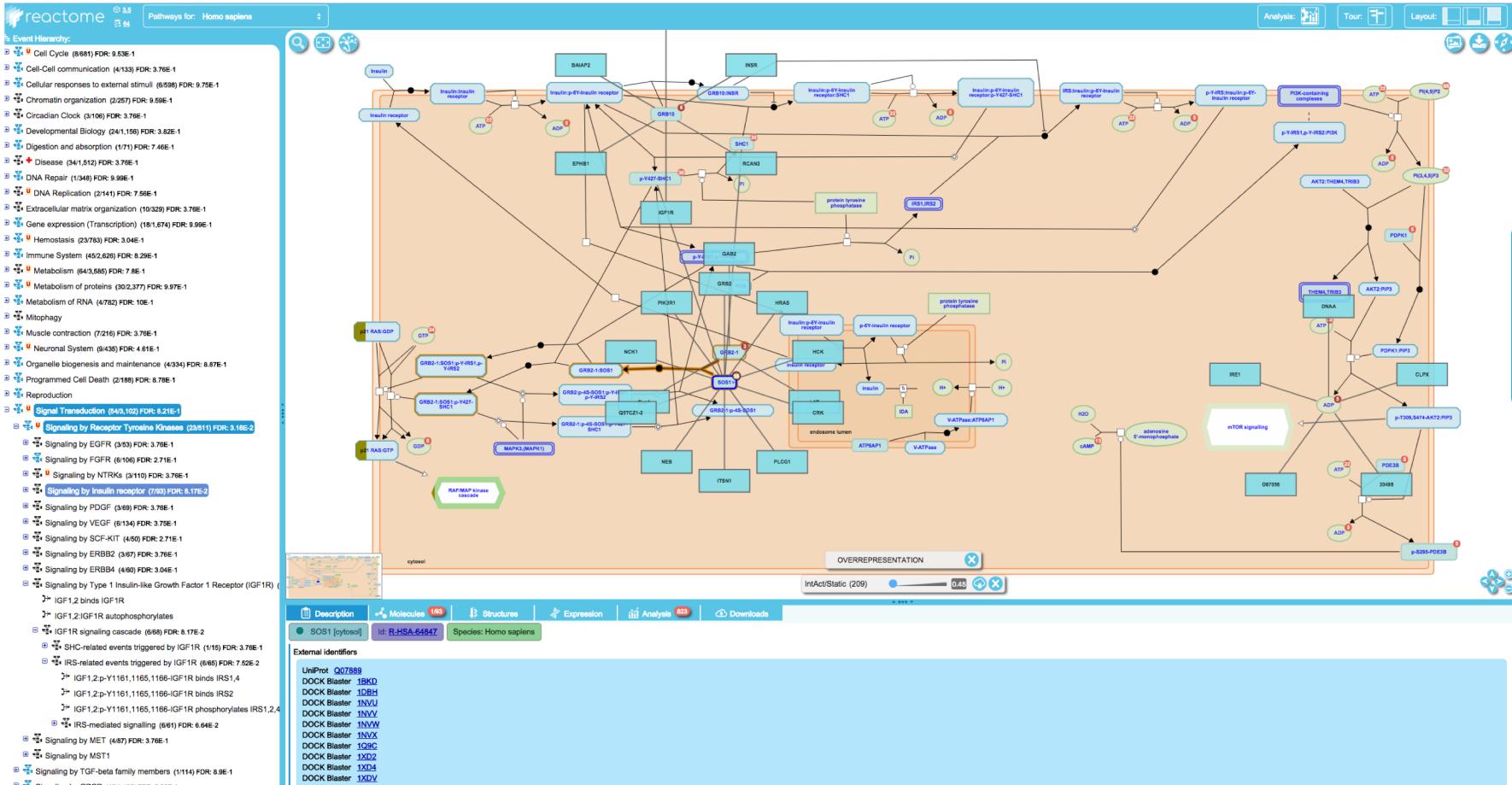
- Cell Cycle (3681) FDR: 9.5E-1
- Cell-Cell communication (4133) FDR: 3.76E-1
- Cellular responses to external stimuli (6598) FDR: 9.75E-1
- Chromatin organization (2257) FDR: 9.59E-1
- Circadian Clock (3106) FDR: 3.76E-1
- Developmental Biology (241,156) FDR: 3.82E-1
- Digestion and absorption (171) FDR: 7.46E-1
- Disease (341,512) FDR: 3.76E-1
- DNA Repair (346) FDR: 9.99E-1
- DNA Replication (2141) FDR: 7.59E-1
- Extracellular matrix organization (10329) FDR: 3.76E-1
- Gene expression (Transcription) (161,674) FDR: 9.99E-1
- Hemostasis (23783) FDR: 3.04E-1
- Immune System (49,2626) FDR: 8.29E-1
- Metabolism (64,3585) FDR: 7.8E-1
- Metabolism of proteins (902,377) FDR: 9.97E-1
- Metabolism of RNA (4782) FDR: 10E-1
- Mitophagy
- Muscle contraction (7216) FDR: 3.79E-1
- Neuronal System (6435) FDR: 4.61E-1
- Organelle biogenesis and maintenance (4334) FDR: 8.87E-1
- Programmed Cell Death (2186) FDR: 8.78E-1
- Reproduction
- Signal Transduction (643,102) FDR: 8.21E-1**
- Signaling by Receptor Tyrosine Kinases (23811) FDR: 3.16E-2**
 - Signaling by EGFR (935) FDR: 3.78E-1
 - Signaling by FGFR (6106) FDR: 2.71E-1
 - Signaling by NTRKs (3119) FDR: 3.76E-1
 - Signaling by Insulin receptor (793) FDR: 8.17E-2**
 - Signaling by PDGF (369) FDR: 3.78E-1
 - Signaling by VEGF (6134) FDR: 3.78E-1
 - Signaling by SCF-KIT (450) FDR: 2.71E-1
 - Signaling by ERBB2 (367) FDR: 3.76E-1
 - Signaling by ERBB4 (460) FDR: 3.04E-1
 - Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R)
 - IGF1.2 binds IGF1R
 - IGF1.2/IGF1R autophosphorylates
 - IGF1R signaling cascade (668) FDR: 8.17E-2
 - SHC-related events triggered by IGF1R (1115) FDR: 3.76E-1
 - IRS-related events triggered by IGF1R (665) FDR: 7.52E-2
 - IGF1.2-p-Y1161,1165-IGF1R binds IRS1.4
 - IGF1.2-p-Y1161,1165-IGF1R binds IRS2

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
IRS-related events triggered by IGF1R	6	65	0.005	1.79E-3	7.52E-2	3	12	0.001	Homo sapiens
Negative regulation of MAPK pathway	5	45	0.003	1.95E-3	7.79E-2	8	12	0.001	Homo sapiens
IGF1R signaling cascade	6	68	0.005	2.24E-3	8.17E-2	4	17	0.001	Homo sapiens
Transport of organic anions	4	28	0.002	2.26E-3	8.17E-2	4	8	0.001	Homo sapiens
Signaling by Insulin receptor	7	83	0.007	2.4E-3	8.17E-2	7	34	0.003	Homo sapiens
Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R)	6	69	0.005	2.4E-3	8.17E-2	4	19	0.002	Homo sapiens
PPARA activates gene expression	10	178	0.013	2.67E-3	8.8E-2	41	41	0.004	Homo sapiens

Reactome: Pathway Details



Reactome: Pathway Details



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

9,559,001 lists analyzed

238,875 terms

129 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.

[Transcription](#)[Pathways](#)[Ontologies](#)[Disease/Drugs](#)[Cell Types](#)[Misc](#)[Legacy](#)[Crowd](#)**Description** No description available (65 genes)**KEGG 2016****WikiPathways 2016****ARCHS4 Kinases Coexp**

GABAergic synapse_Homo sapiens_hsa0472
Morphine addiction_Homo sapiens_hsa0503
Glutamatergic synapse_Homo sapiens_hsa0...
Dopaminergic synapse_Homo sapiens_hsa0...
Circadian entrainment_Homo sapiens_hsa0...

Calcium Regulation in the Cardiac Cell_Hom...
Calcium Regulation in the Cardiac Cell_Mus...
Myometrial Relaxation and Contraction Path...
Myometrial Relaxation and Contraction Path...
G Protein Signaling Pathways_Mus musculus

CAMKV_human_kinase_ARCHS4_coexpressio...
PNCK_human_kinase_ARCHS4_coexpression
CAMK1G_human_kinase_ARCHS4_coexpress...
CAMK2B_human_kinase_ARCHS4_coexpress...
CAMK4_human_kinase_ARCHS4_coexpressio...

Reactome 2016**BioCarta 2016****HumanCyc 2016**

Neuronal System_Homo sapiens_R-HSA-112...
Transmission across Chemical Synapses_Ho...
Glucagon signaling in metabolic regulation_I...
Vasopressin regulates renal water homeostat...
Opioid Signaling_Homo sapiens_R-HSA-1111

Regulation of ck1/cdk5 by type 1 glutamate...
GATA3 participate in activating the Th2 cytol...
How Progesterone Initiates the Oocyte Matu...
Cystic fibrosis transmembrane conductance...
Activation of cAMP-dependent protein kinas...

GABA shunt_Homo sapiens_GLUEG-I-PWY
Valine degradation_Homo sapiens_VALDEG-I

NCI-Nature 2016**Panther 2016****BioPlex 2017**

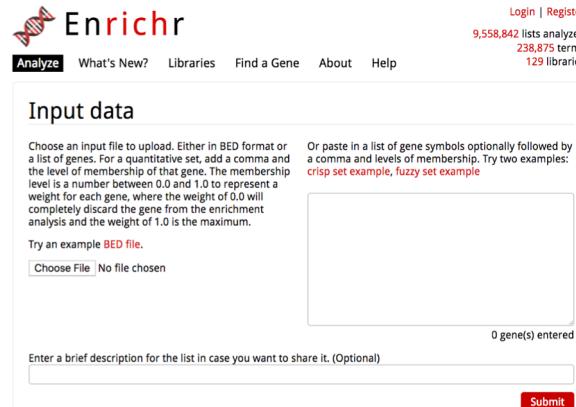
Endothelins_Homo sapiens_dfb9dc47-6191...
LPA4-mediated signaling events_Homo sapi...
LPA receptor mediated events_Homo sapien...
Plasma membrane estrogen receptor signal...
IL8- and CXCR2-mediated signaling events_H...

GABA-B receptor II signaling_Homo sapien...
Heterotrimeric G-protein signaling pathway...
Heterotrimeric G-protein signaling pathway...
Heterotrimeric G-protein signaling pathway...
Endothelin signaling pathway_Homo sapien...

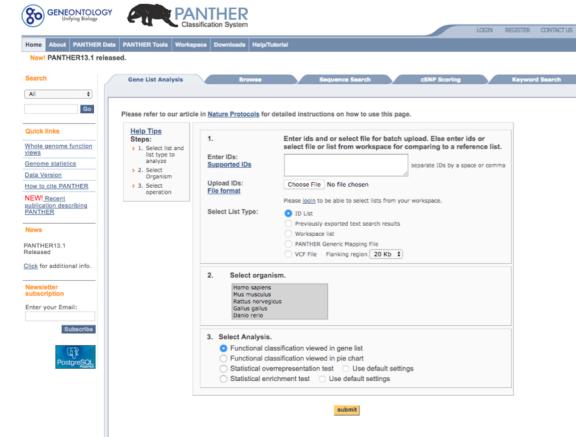
C14orf104
NUFIP1
ARMC1
HAX1
DGUOK

Gene Ontology

- Enrichr
 - <http://amp.pharm.mssm.edu/Enrichr/>
- Panther
 - <http://www.pantherdb.org/>



The screenshot shows the Enrichr web interface. At the top right, there are links for 'Login | Register' and statistics: '9,558,842 lists analyzed', '238,875 terms', and '129 libraries'. Below this is a navigation bar with 'Analyze', 'What's New?', 'Libraries', 'Find a Gene', 'About', and 'Help'. The main area is titled 'Input data' with a sub-instruction: 'Choose an input file to upload. Either in BED format or a list of genes. For a quantitative score add a comma and the membership level 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.' A red link 'Try an example BED file.' is present. There is a 'Choose File' button with the message 'No file chosen'. To the right is a text input field containing '0 gene(s) entered' and a 'Submit' button. Below this is a 'Enter a brief description for the list in case you want to share it. (Optional)' input field and a 'Contribute' checkbox.



The screenshot shows the PANTHER Classification System homepage. At the top, there are links for 'Home', 'About', 'PANTHER Data', 'PANTHER Tools', 'Workspaces', 'Downloads', 'Help/Tutorials', 'LOGON', 'REGISTER', and 'CONTACT US'. A message 'New! PANTHER13.1 released!' is displayed. The main content area has tabs for 'Search', 'Gene List Analysis', 'Browse', 'Sequence Search', 'cIMP Scoring', and 'Keyword Search'. On the left, there is a sidebar with 'Quick Links' including 'Whole genome function', 'Data', 'Genome statistics', 'Data Version', 'How to use PANTHER', 'NEW! Recent publications', 'PANTHER', 'News', 'PANTHER13.1 Released', 'How to cite PANTHER', 'Newsletter', 'Subscription', and 'Enter your Email'. The 'Gene List Analysis' tab is active. It contains a form with steps: 1. Enter IDs or select file for batch upload, 2. Select organism, 3. Select annotation. It also includes sections for 'Select List Type' (with options for ID List, Pantherically exported text search results, Workspace list, PANTHER Generic Mapping File, and VCF File), 'Select organism' (with options for Homo sapiens, Mus musculus, Rattus norvegicus, Gallus gallus, Danio rerio), and 'Select Analysis' (with options for Functional classification viewed in list, Functional classification viewed in pie chart, Statistical overrepresentation test, and Statistical enrichment test). A 'Submit' button is at the bottom right.

[Login](#) | [Register](#)[Transcription](#) [Pathways](#) [Ontologies](#) [Disease/Drugs](#) [Cell Types](#) [Misc](#) [Legacy](#) [Crowd](#)**Description** No description available (65 genes)**GO Cellular Component** [i](#)
2017b

ionotropic glutamate receptor complex (GO:0005524)
clathrin-sculpted gamma-aminobutyric acid
dendritic filopodium (GO:1902737)
alphav-beta5 integrin-vitronectin complex (GO:0005525)
apical plasma membrane urothelial plaque (GO:0005526)

GO Biological Process [i](#)
2017b

cellular response to glucagon stimulus (GO:0005527)
synaptic transmission, GABAergic (GO:00519)
regulation of calcium:sodium antiporter activity (GO:0005528)
regulation of sodium:proton antiporter activity (GO:0005529)
regulation of taurine selective glutamate receptor activity (GO:0005530)

GO Molecular Function [i](#)
2017b

GTPase motor activity (GO:0061791)
GTPase activity, coupled (GO:0061745)
GTPase activity (GO:0003924)
protein serine/threonine phosphatase inhibitor (GO:0004165)
myosin V binding (GO:0031489)

MGI Mammalian Phenotype [i](#)
2017

MP:0002064_seizures
MP:0001402_hypoactivity
MP:0004924_abnormal_behavior
MP:0002169_no_abnormal_phenotype_detected
MP:0001399_hyperactivity

Human Phenotype Ontology [i](#)

Depression (HP:0000716)
Anxiety (HP:0000739)
Heart block (HP:0012722)
Febrile seizures (HP:0002373)
Atonic seizures (HP:0010819)

Jensen TISSUES [i](#)

Parietal_lobe
Frontal_lobe
Occipital_lobe
Caudate_nucleus
Pons

Jensen COMPARTMENTS [i](#)

Neuron_part
Synapse
Cell_projection
Neuron_projection
membrane-bounded_vesicle

Jensen DISEASES [i](#)

Succinic_semialdehyde_dehydrogenase_deficiency
Stiff-Person_syndrome
Lung_cancer
Adenoma
Diamond-Blackfan_anemia

Panther: Submitting Gene List

The screenshot shows the PANTHER Classification System homepage with a navigation bar at the top. The main content area is titled 'Gene List Analysis' and contains the following steps:

- Step 1: Enter IDs**
A text input field contains the IDs: Abat, Abim2, Adcy1. A note says "separate IDs by a space or comma".
Below it is a "Choose File" button and a message "No file chosen".
A "Select List Type:" section includes radio buttons for "ID List" (selected), "Previously exported text search results", "Workspace list", "PANTHER Generic Mapping File", and "VCF File Flanking region 20 Kb".
- Step 2: Select organism.**
A dropdown menu is open, showing species names: Homo sapiens, Mus musculus (selected), Rattus norvegicus, Gallus gallus, and Danio rerio.
- Step 3: Select Analysis.**
Radio buttons for analysis type:
 - Functional classification viewed in gene list
 - Functional classification viewed in pie chart** (selected)
 - Statistical overrepresentation test
 - Statistical enrichment test

A "submit" button is located at the bottom right of the form.

Annotations:

- A green callout box on the right says "Paste Gene list" with an arrow pointing to the input field.
- A green callout box below it says "Choose Species E.g. Mouse" with an arrow pointing to the dropdown menu.
- A green callout box at the bottom right says "Select Analysis Type e.g. Pie Chart" with an arrow pointing to the "Functional classification viewed in pie chart" radio button.

At the bottom of the page, there is a footer with links: About | Release Information | Contact Us | System Requirements | Privacy Policy | Disclaimer. © Copyright 2018 Paul Thomas All Rights Reserved.

Panther



GENEONTOLOGY
Unifying Biology



LOGIN REGISTER CONTACT US

Home About PANTHER Data PANTHER Tools Workspace Downloads Help/Tutorial

New! PANTHER13.1 released.



Bar Chart Export

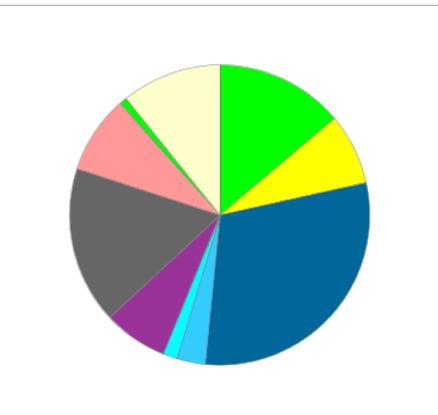
Features:

- Mouse-over pie chart section to see category name and statistics
- Click on a pie chart section to drill down to child categories
- Click on chart legend link to retrieve gene list for each category
- Click on a color key in chart legend to choose your favorite color for the category NEW!
- Click on in chart legend to highlight your selection in pie chart NEW!
- Click on to reset

Select Ontology: Biological Process View: 100%

PANTHER GO-Slim Biological Process

Total # Genes: 67 Total # process hits: 130



Click to get gene list for a category:

- █ biological regulation (GO:0065007)
- █ cellular component organization or biogenesis (GO:0071840)
- █ cellular process (GO:0009987)
- █ developmental process (GO:0032502)
- █ growth (GO:0040007)
- █ localization (GO:0051179)
- █ metabolic process (GO:0008152)
- █ multicellular organismal process (GO:0032501)
- █ reproduction (GO:0000003)
- █ response to stimulus (GO:0050896)

Color picker powered by

**Chart tooltips are read as: Category name (Accession): # genes; Percent of gene hit against total # genes; Percent of gene hit against total # Process hits

Panther



GENEONTOLOGY
Unifying Biology



LOGIN REGISTER CONTACT US

Home About PANTHER Data PANTHER Tools Workspace Downloads Help/Tutorial

New! PANTHER13.1 released.

PANTHER PIE CHART

[Bar Chart](#) [Export](#)

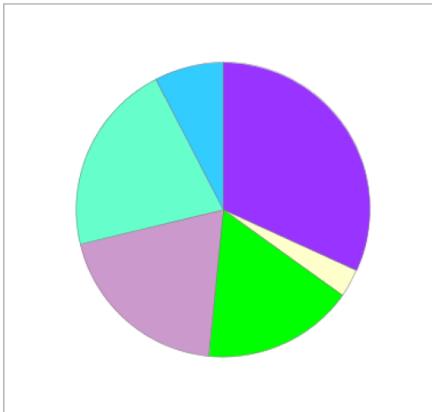
Features:

- Mouse-over pie chart section to see category name and statistics
- Click on a pie chart section to drill down to child categories
- Click on chart legend link to retrieve gene list for each category
- Click on a color key in chart legend to choose your favorite color for the category **NEW!**
- Click on in chart legend to highlight your selection in pie chart **NEW!**
- Click on to reset

Select Ontology: [Cellular Component](#) View: [100%](#)

PANTHER GO-Slim Cellular Component

Total # Genes: 67 Total # component hits: 66



Click to get gene list for a category:

- [cell part \(GO:0044464\)](#)
- [extracellular region \(GO:0005576\)](#)
- [macromolecular complex \(GO:0032991\)](#)
- [membrane \(GO:0016020\)](#)
- [organelle \(GO:0043226\)](#)
- [synapse \(GO:0045202\)](#)

Color picker powered by Web Colors by VisiBone

**Chart tooltips are read as: Category name (Accession): # genes; Percent of gene hit against total # genes; Percent of gene hit against total # Component hits

Panther



GENEONTOLOGY
Unifying Biology



LOGIN REGISTER CONTACT US

Home About PANTHER Data PANTHER Tools Workspace Downloads Help/Tutorial

New! PANTHER13.1 released.

PANTHER PIE CHART

[Bar Chart](#) [Export](#)

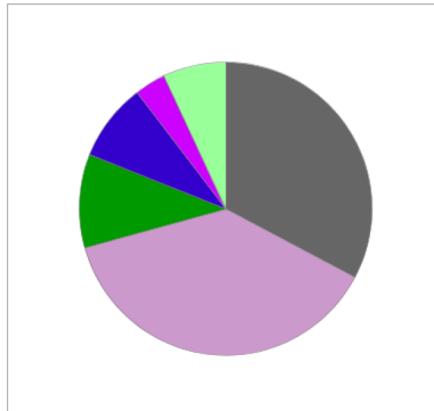
Features:

- > Mouse-over pie chart section to see category name and statistics
- > Click on a pie chart section to drill down to child categories
- > Click on chart legend link to retrieve gene list for each category
- > Click on a color key in chart legend to choose your favorite color for the category NEW!
- > Click on in chart legend to highlight your selection in pie chart NEW!
- > Click on to reset

Select Ontology: [Molecular Function](#) View: [100%](#)

PANTHER GO-Slim Molecular Function

Total # Genes: 67 Total # function hits: 58



Click to get gene list for a category:

- [binding \(GO:0005488\)](#)
- [catalytic activity \(GO:0003824\)](#)
- [receptor activity \(GO:0004872\)](#)
- [signal transducer activity \(GO:0004871\)](#)
- [structural molecule activity \(GO:0005198\)](#)
- [transporter activity \(GO:0005215\)](#)

Color picker powered by VisiBone

**Chart tooltips are read as: Category name (Accession): # genes; Percent of gene hit against total # genes; Percent of gene hit against total # Function hits

Panther

GENEONTOLOGY
Unifying BiologyPANTHER
Classification System

LOGIN REGISTER CONTACT US

Home About PANTHER Data PANTHER Tools Workspace Downloads Help/Tutorial

New! PANTHER13.1 released.

PANTHER GENE LIST Customize Gene list

Convert List to: Display: items per page [Refine Search](#)

Hits 1-5 of 5 [page: (1)]

	Gene ID	Mapped IDs	Gene Name Gene Symbol Ortholog	PANTHER Family/Subfamily	PANTHER Protein Class	Species
<input type="checkbox"/>	1. MOUSE MGI=MGI=1927653 UniProtKB=Q9EPBZ	Gpr88	Probable G-protein coupled receptor 88 Gpr88 ortholog	G-PROTEIN COUPLED RECEPTOR BB-RELATED (PTHR2420;SF143)	G-protein coupled receptor	Mus musculus
<input type="checkbox"/>	2. MOUSE MGI=MGI=3613677 UniProtKB=D3YZU1	Shank1	SH3 and multiple ankyrin repeat domains protein 1 Shank1 ortholog	SH3 AND MULTIPLE ANKYRIN REPEAT DOMAINS PROTEIN 1 (PTHR24135;SF3)	-	Mus musculus
<input type="checkbox"/>	3. MOUSE MGI=MGI=95774 UniProtKB=QBCGK7	Gna1	Guanine-nucleotide-binding protein G(o/l) subunit alpha Gna1 ortholog	GUANINE NUCLEOTIDE-BINDING PROTEIN (GOLF) SUBUNIT ALPHA (PTHR10218;SF233)	heterotrimeric G-protein	Mus musculus
<input type="checkbox"/>	4. MOUSE MGI=MGI=95777 UniProtKB=Q6R0HZ	Gnas	Guanine-nucleotide-binding protein G(s) subunit alpha isoforms Xlas Xlas Gnas ortholog	GUANINE NUCLEOTIDE-BINDING PROTEIN G(S) SUBUNIT ALPHA ISOFORMS XLAS (PTHR10218;SF212)	heterotrimeric G-protein	Mus musculus
<input type="checkbox"/>	5. MOUSE MGI=MGI=95613 UniProtKB=P62812	Gabra1	Gamma-aminobutyric acid receptor alpha-1 Gabra1 ortholog	GAMMA-AMINOBUTYRIC ACID RECEPTOR SUBUNIT ALPHA-1 (PTHR18945;SF514)	GABA receptor acetylcholine receptor	Mus musculus

Hits 1-5 of 5 [page: (1)]

R

- `source("https://bioconductor.org/biocLite.R")`
- `biocLite("AnnotationDbi")`
- `biocLite("org.Mm.eg.db")`
- `biocLite("edgeR")`

R Goana

```
library("AnnotationDbi")
library("org.Mm.eg.db")
library(edgeR) ## For goana

## Read GeneList file
DIR <- "/Users/kumarpa/Desktop/Work/KUL/MyPub/Presentations/Jax/Others"
setwd(DIR)
DGEs <- read.table(paste0("GeneList.txt"), header=T, as.is=T); head(DGEs); dim(DGEs)
rownames(DGEs) <- DGEs$Genes; head(DGEs); dim(DGEs)

## Map GeneIDs
DGEs$entrez = mapIds(org.Mm.eg.db, keys=rownames(DGEs), column="ENTREZID", keytype="SYMBOL", multiVals="first"); head(DGEs)

## GO using Goana
go <- goana(as.vector(DGEs$entrez), geneid = rownames(DGEs), species="Mm")
GoanaBP <- topGO(go, n=20, ontology = c("BP")); GoanaBP

GoanaMF <- topGO(go, n=20, ontology = c("MF")); GoanaMF

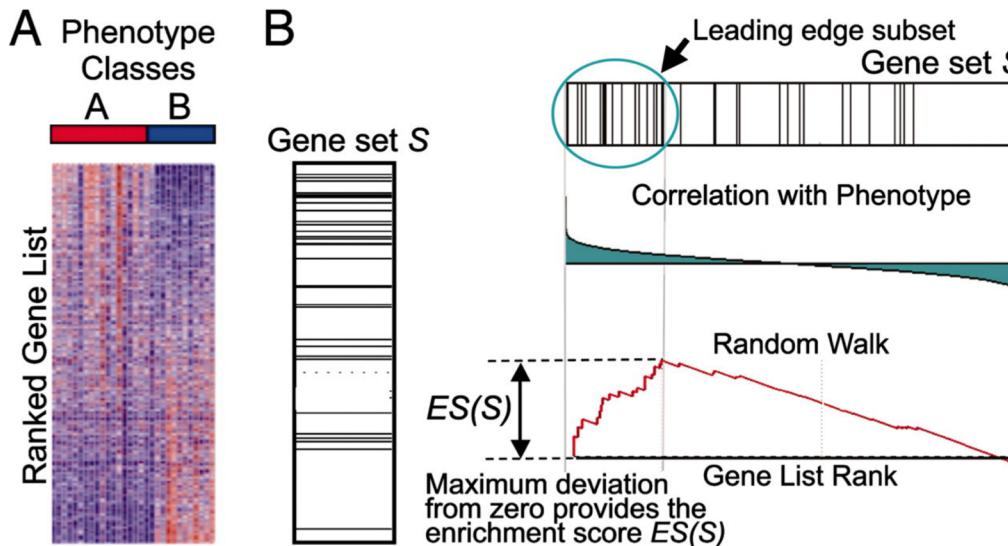
GoanaCC <- topGO(go, n=20, ontology = c("CC")); GoanaCC
write.table(GoanaCC, file=paste0("Goana.GO_CC_for_DGEsOnly.Top20.txt"), row.names=T, quote=F, sep="\t")

keg <- kegg(as.vector(DGEs$entrez), species="Mm")
GoanaKEGG <- topKEGG(keg, n=20, truncate=20); GoanaKEGG
write.table(GoanaKEGG, file=paste0("Goana.GO_KEGG_for_DGEsOnly.Top20.txt"), row.names=T, quote=F, sep="\t")
```

Gene Set Enrichment Analysis

- determines whether pre-defined set of genes shows statistically significant, concordant differences between two biological states
- resolves the problem of the undetectable, small changes in the expression of single genes

Schematic overview of GSEA



Gene Set Database File

A geneset database (.gmt) file is a tab separated text file containing one geneset per line.

The first column is the gene set name.

The second column is a brief description of the gene set.

The remaining columns contain the names of the genes in the gene set. Notes:

Sample Phenotype File

A sample phenotype (.cls) file is a text file containing three lines.

The first line contains three numbers separated by spaces.
The first number is the number of samples. The second and third numbers are the constants 2 and 1, respectively.

The second line begins with # and is followed by a space separated list of “long” phenotype names.

The third line consists of a space separated list of “short” phenotype labels for each of the samples in the gene expression file, in the same order they occur there.

Gene Expression File

A gene expression (.gct) file is a tab-separated text file.

The first line is the constant #1.2.

The second line contains two numbers: the number of genes and the number of samples.

The third line contains column headers for the table on the following lines. The fourth line and below each contain expression values for one gene:

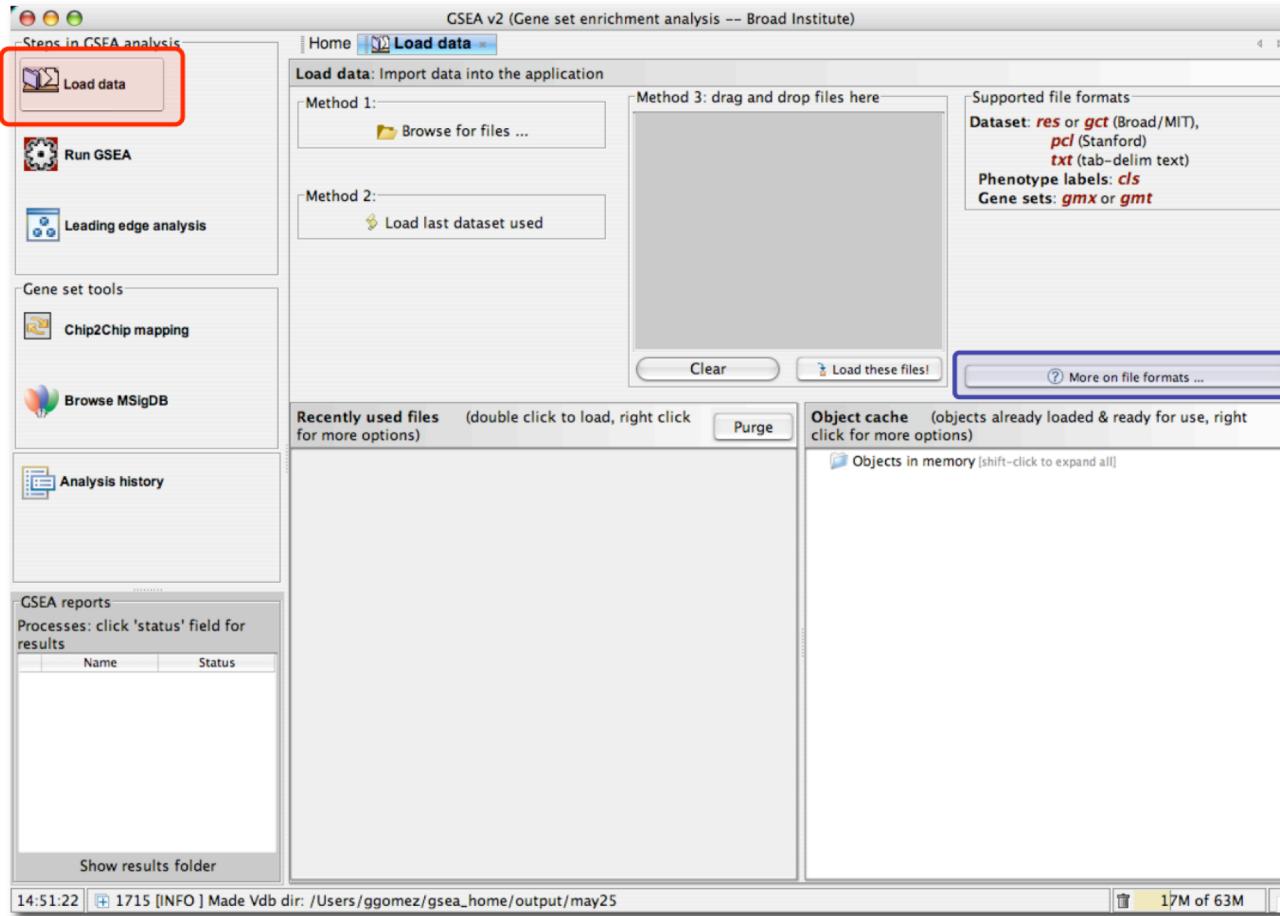
- the first column is the gene name,
- the second column is the gene description,
- the third and subsequent columns are the expression values for each sample.

Annotation file (.chip)

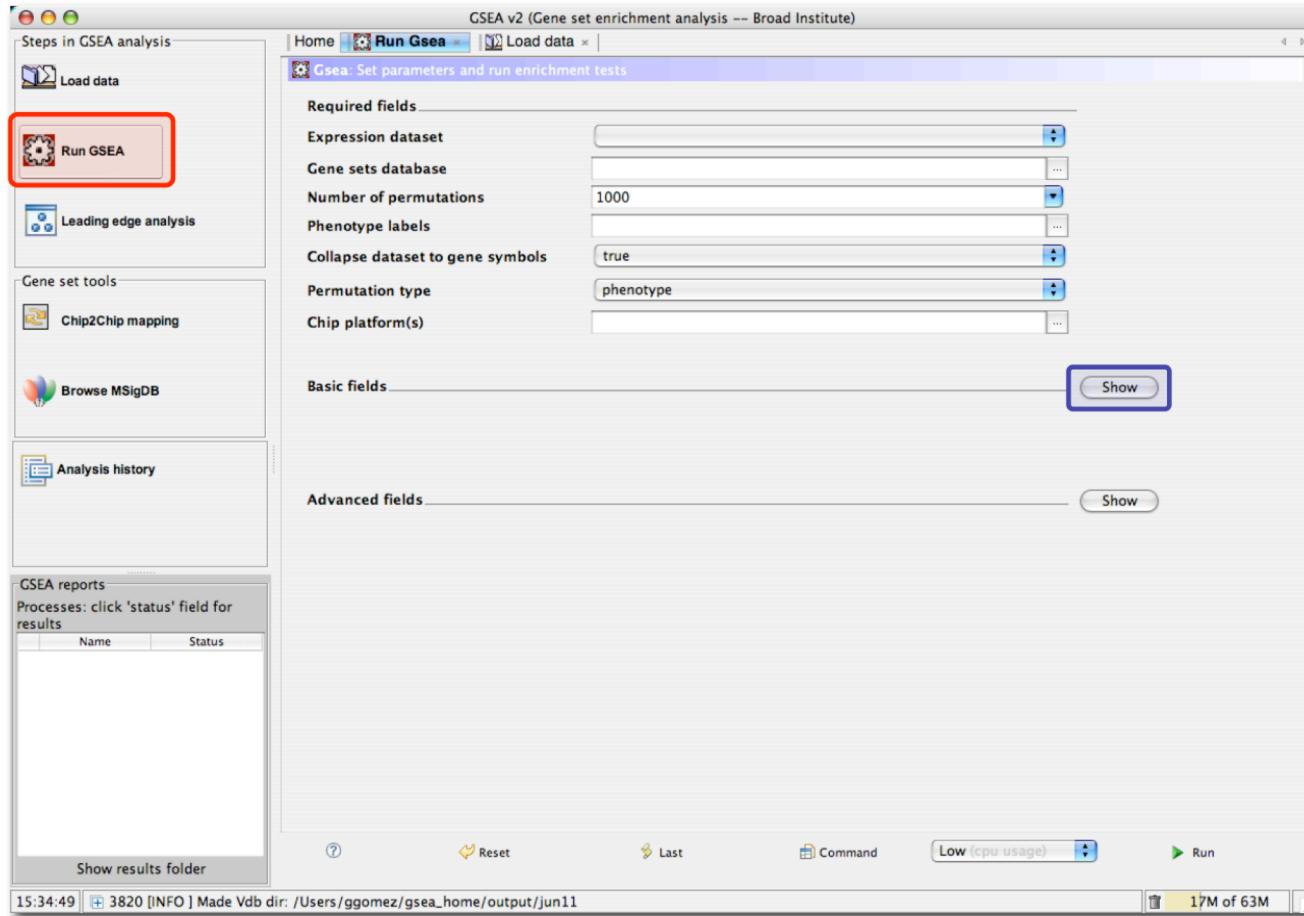
The first line contains column headings that identify the content of each column in the remainder of the file. The file must contain three column headings separated by tabs:

- Probe Set ID
- Gene Symbol
- Gene Title

GSEA: Loading Data



Running GSEA



How to interpret results

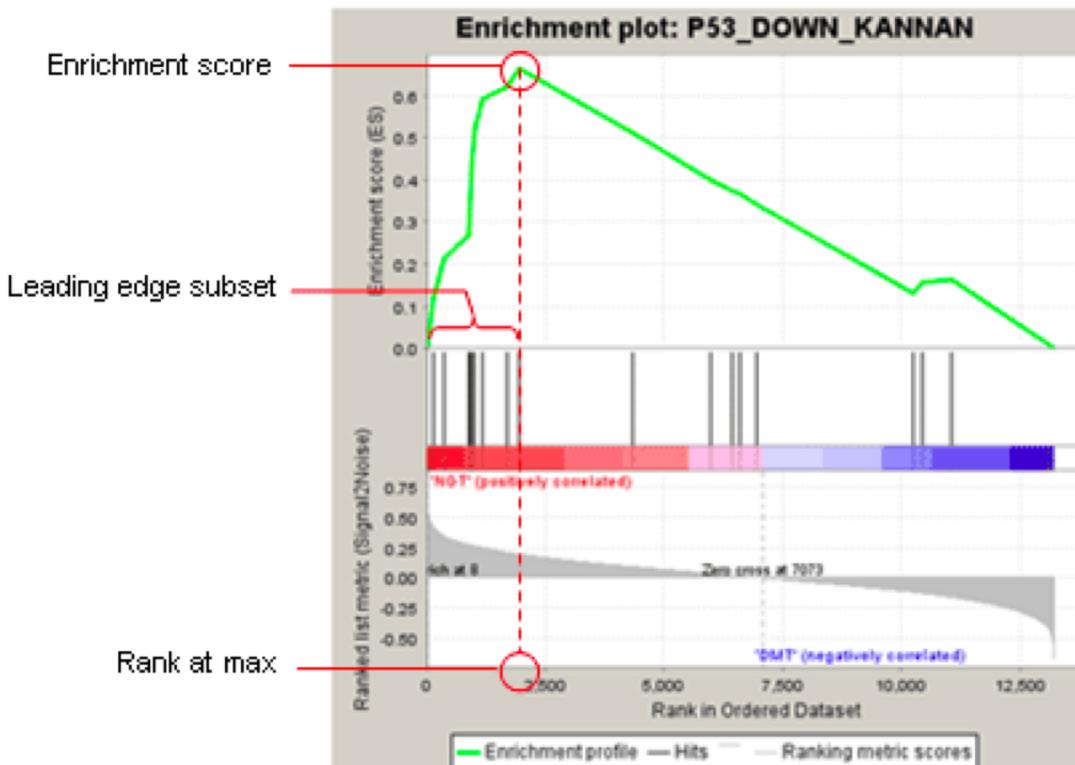


Fig 1: Enrichment plot: P53_DOWN_KANNAN
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

Thanks