

Gene set enrichment analysis

Genomics sub-group meeting 3

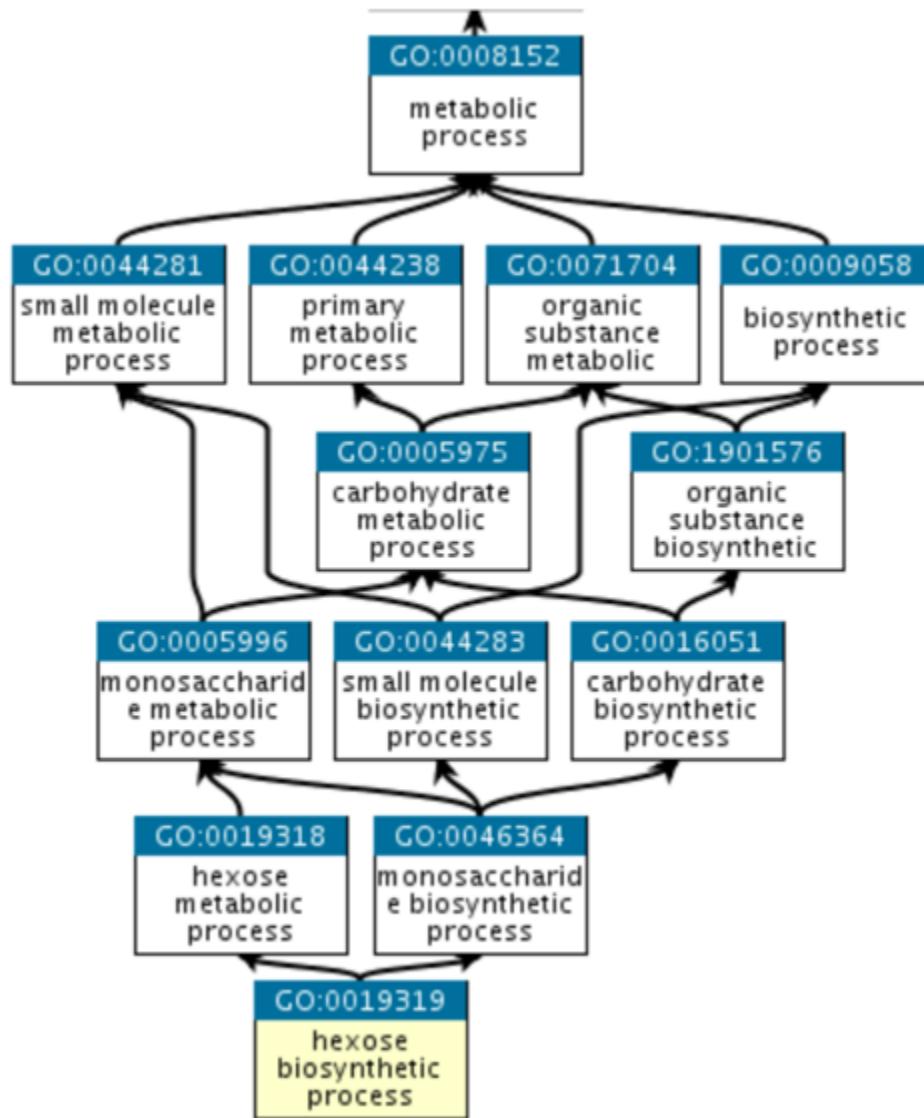
2020/02/17

Jon Sánchez Valle

Outline

- DAVID
- g:Profiler
- ToppGene
- GSEA
- Other methods + visualization

Gene Ontology



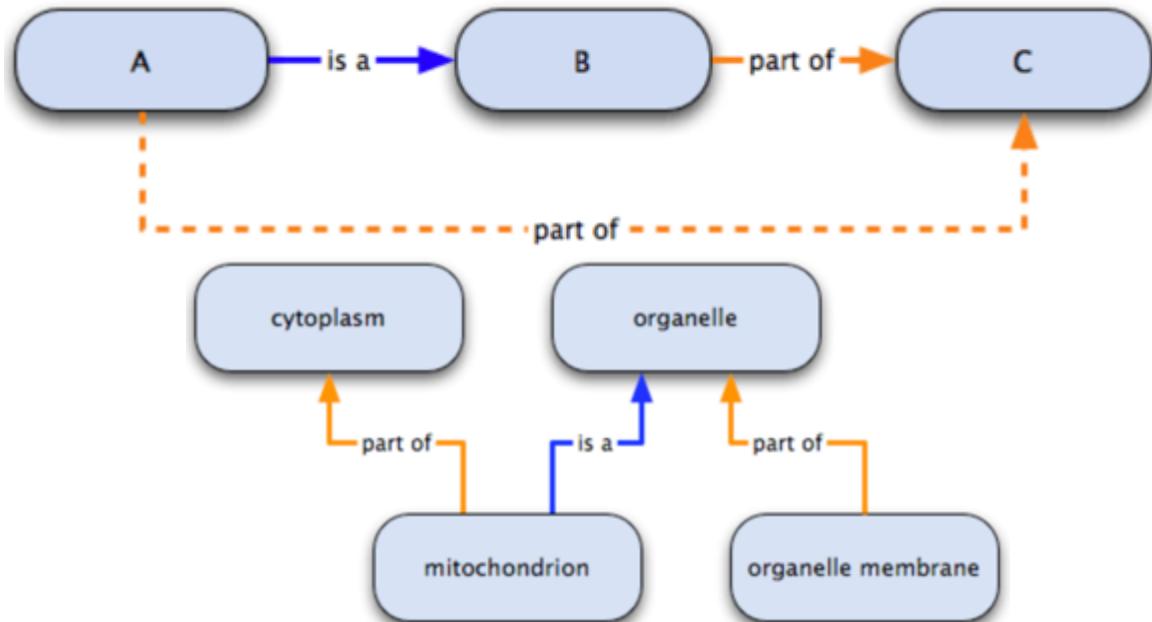
GO:1904659:glucose transport is a GO:0015749:monosaccharide transport.

The Gene Ontology employs a number of other relations, including *part of*, e.g.

GO:0031966:mitochondrial membrane is *part of* GO:0005740:mitochondrial envelope

and *regulates*, e.g:

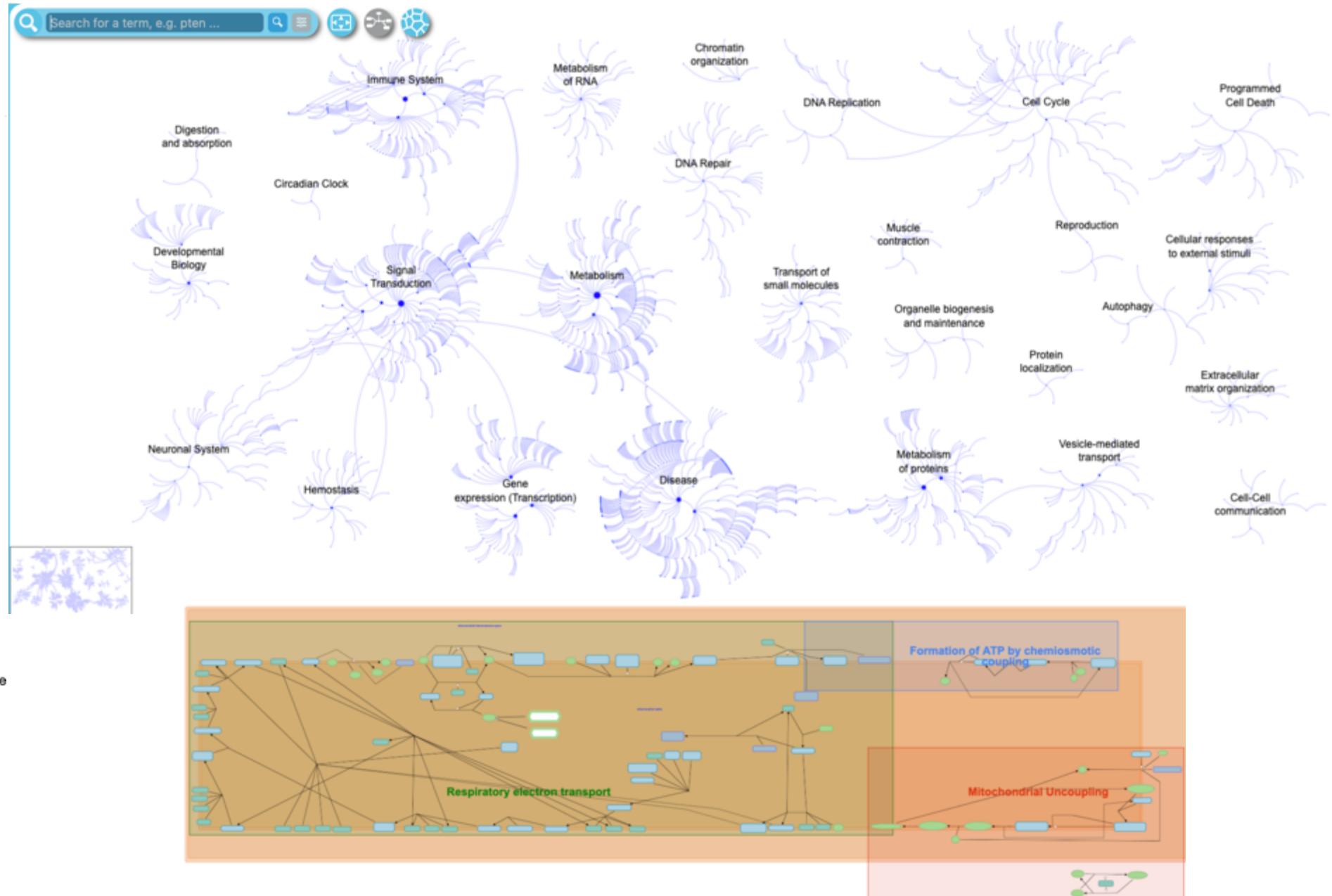
GO:0098689:latency-replication decision *regulates* GO:0019046:release from viral latency



Reactome



- Autophagy
- Cell Cycle
- Cell-Cell communication
- Cellular responses to external stimuli
- Chromatin organization
- Circadian Clock
- Developmental Biology
- Developmental Biology
- Digestion and absorption
- Disease
- DNA Repair
- DNA Replication
- Extracellular matrix organization
- Gene expression (Transcription)
- Hemostasis
- Immune System
- Metabolism
- Metabolism of proteins
- Metabolism of RNA
- Muscle contraction
- Neuronal System
- Organelle biogenesis and maintenance
- Programmed Cell Death
- Protein localization
- Reproduction
- Signal Transduction
- Transport of small molecules
- Vesicle-mediated transport



DAVID (Database for Annotation, Visualization, and Integrated Discovery)

Genome Biology

Software

Open Access

DAVID: Database for Annotation, Visualization, and Integrated Discovery

Glynn Dennis Jr*, Brad T Sherman*, Douglas A Hosack*, Jun Yang*,
Wei Gao*, H Clifford Lane[†] and Richard A Lempicki*

14 August, 2003 4,145 citations

DAVID

Expedites functional annotation and analysis of genes encoded by human, mouse, rat or fly genomes

Download weekly data from:

Table I

Sources of annotation data integrated into DAVID

Resource	URL	Reference
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html	[25]
UniGene	http://www.ncbi.nlm.nih.gov/UniGene/	[26]
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/	[27]
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/	[28]
KEGG	http://www.genome.ad.jp/kegg/	[29]
OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM	[30]
Gene Ontology	http://www.geneontology.org/	[31]
University of Michigan	http://dot.ped.med.umich.edu:2000/ourimage/pub/shared/JMR_pub_affyannot.html	[11]
NetAffx	http://www.affymetrix.com/analysis/index.affx	[12]

Annotation Tool

*Gene symbol
GenBank
OMIM
Affymetrix*

DomainCharts module

*PFAM
protein
domains*

Domain designations are linked to the Conserved Domain Database where details regarding domain function, structure and sequence are available

GoCharts module

*Biological process
Molecular function
Cellular component*

Graphically display the distribution of DEG among functional categories

KeggCharts module

*KEGG
biochemical
pathways*

Gives you the number of genes by category

DAVID

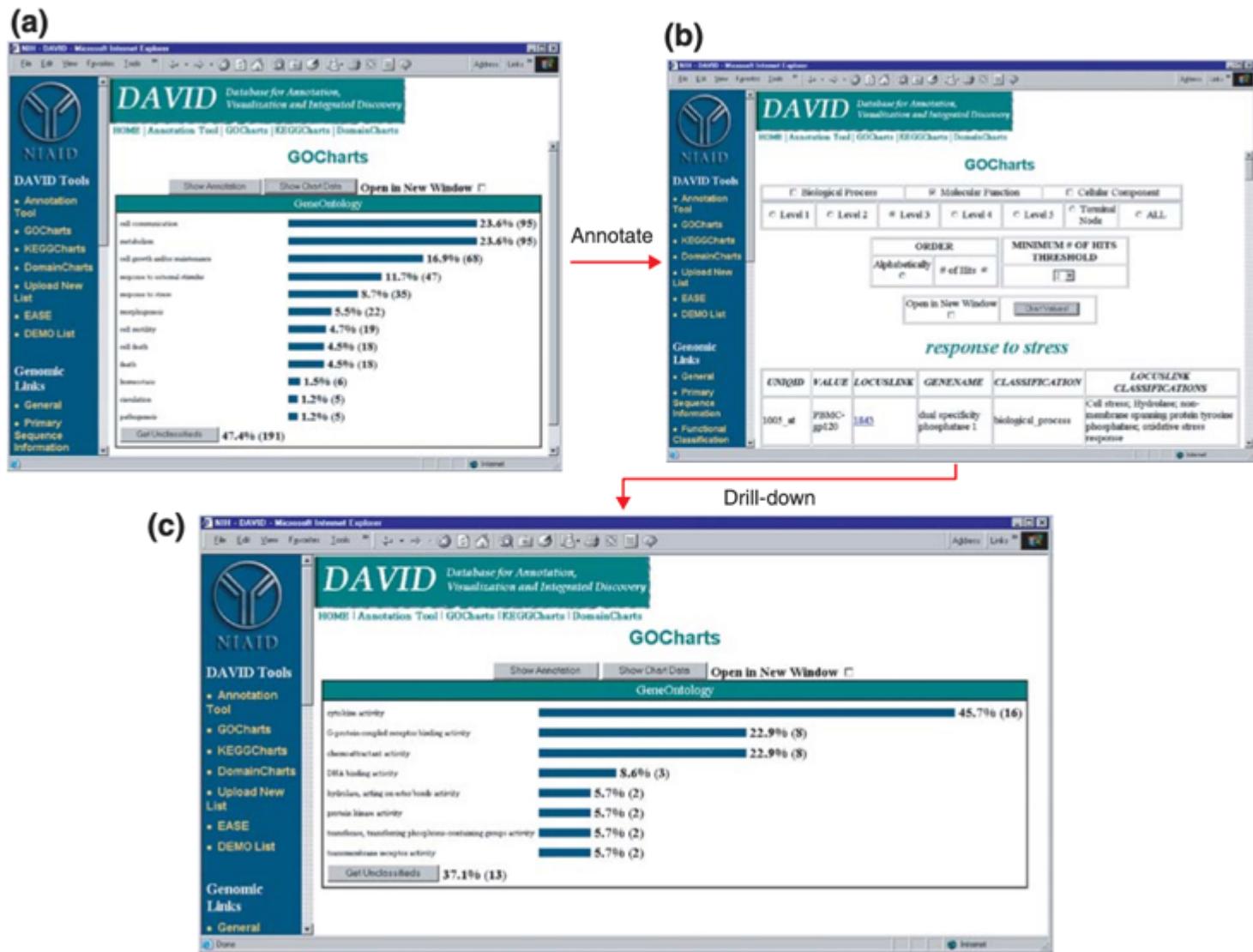
Expedites functional annotation and analysis of genes encoded by human, mouse, rat or fly genomes

Download weekly data from:

Table I

Sources of annotation data integrated into DAVID

Resource	URL	Reference
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html	[25]
UniGene	http://www.ncbi.nlm.nih.gov/UniGene/	[26]
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/	[27]
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/	[28]
KEGG	http://www.genome.ad.jp/kegg/	[29]
OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM	[30]
Gene Ontology	http://www.geneontology.org/	[31]
University of Michigan	http://dot.ped.med.umich.edu:2000/ourimage/pub/shared/JMR_pub_affyannot.html	[11]
NetAffx	http://www.affymetrix.com/analysis/index.affx	[12]



DAVID

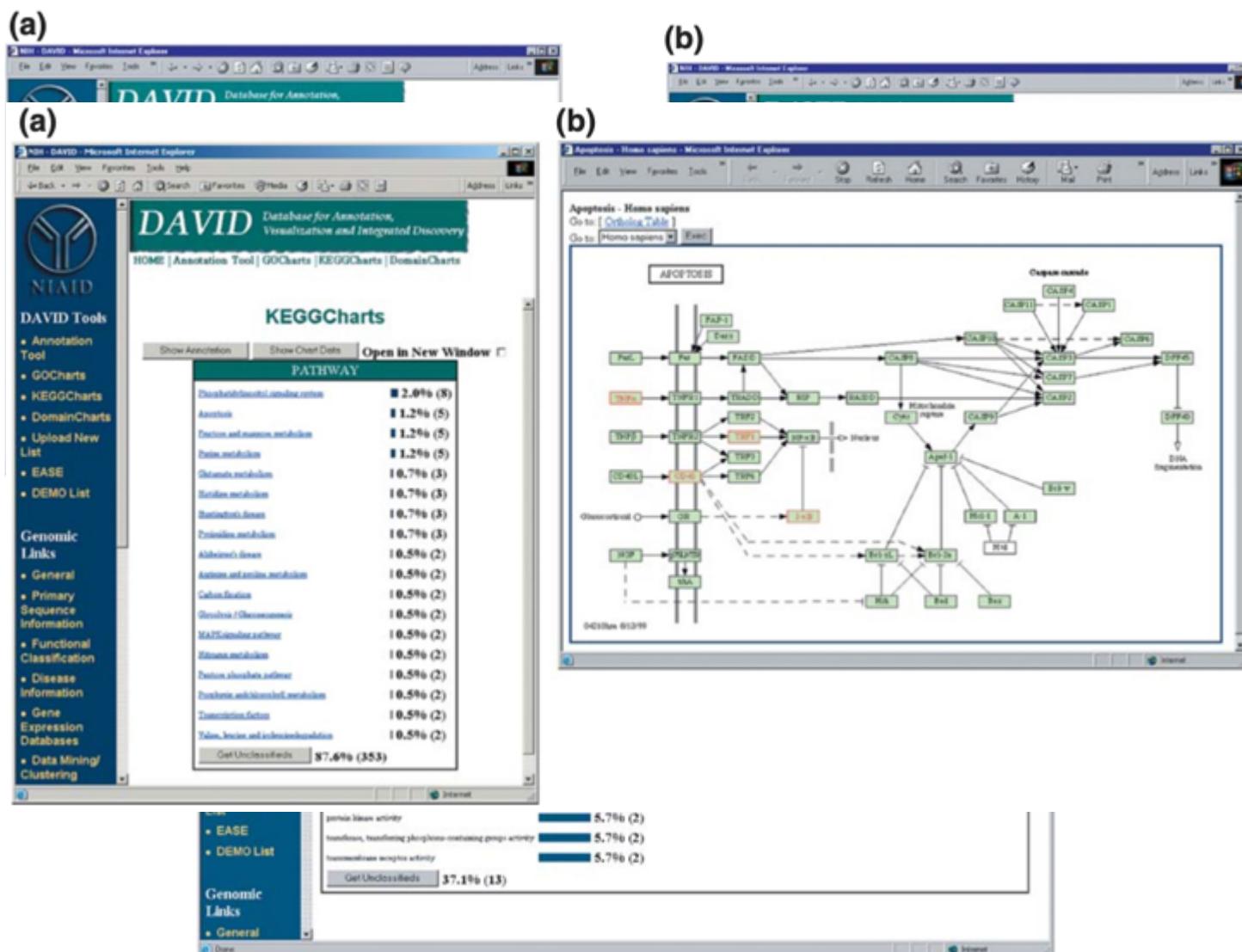
Expedites functional annotation and analysis of genes encoded by human, mouse, rat or fly genomes

Download weekly data from:

Table I

Sources of annotation data integrated into DAVID

Resource	URL	Reference
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html	[25]
UniGene	http://www.ncbi.nlm.nih.gov/UniGene/	[26]
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/	[27]
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/	[28]
KEGG	http://www.genome.ad.jp/kegg/	[29]
OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM	[30]
Gene Ontology	http://www.geneontology.org/	[31]
University of Michigan	http://dot.ped.med.umich.edu:2000/ourimage/pub/shared/JMR_pub_affyannot.html	[11]
NetAffx	http://www.affymetrix.com/analysis/index.affx	[12]



DAVID

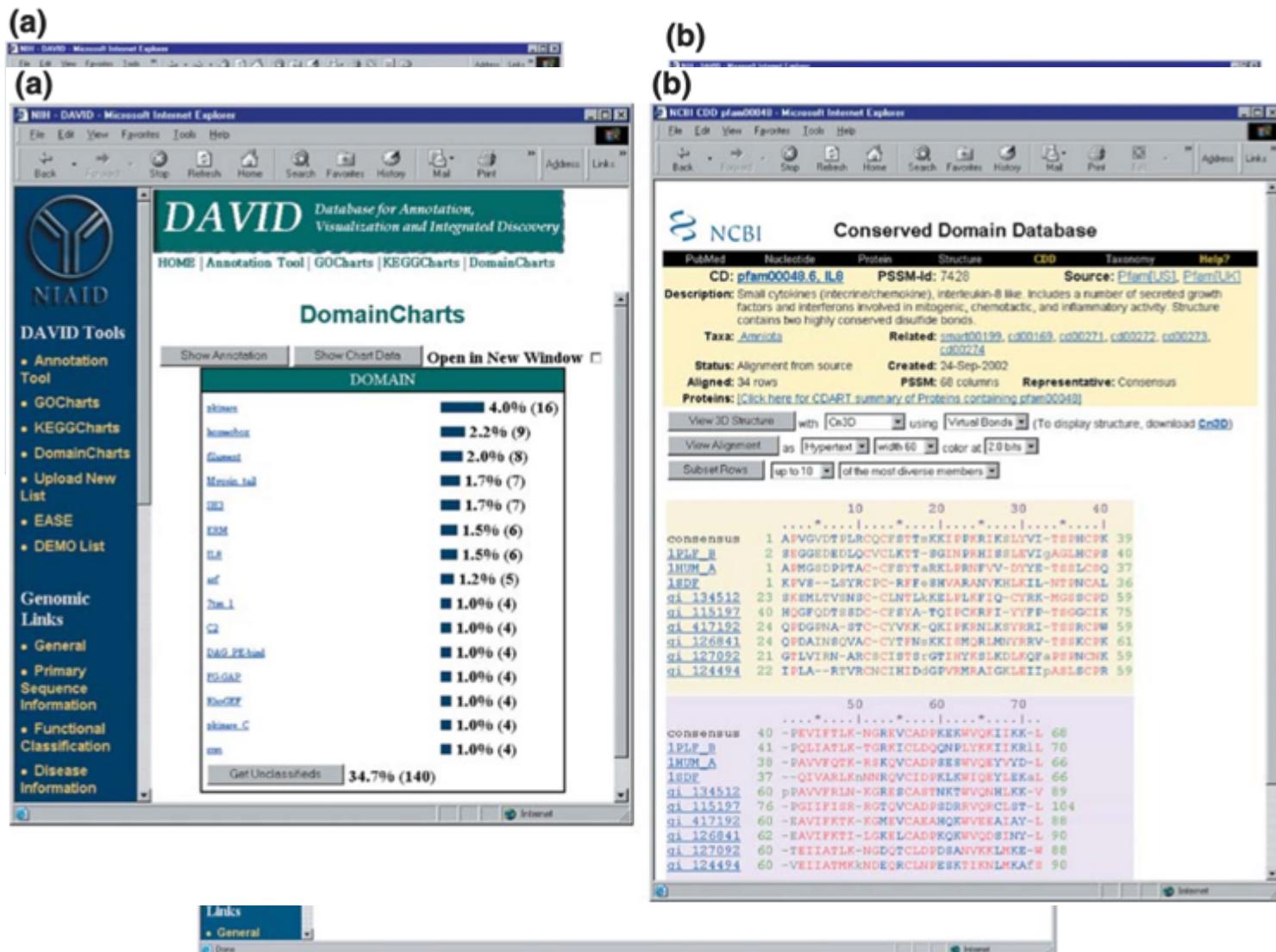
Expedites functional annotation and analysis of genes encoded by human, mouse, rat or fly genomes

Download weekly data from:

Table I

Sources of annotation data integrated into DAVID

Resource	URL	Reference
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html	[25]
UniGene	http://www.ncbi.nlm.nih.gov/UniGene/	[26]
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/	[27]
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/	[28]
KEGG	http://www.genome.ad.jp/kegg/	[29]
OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM	[30]
Gene Ontology	http://www.geneontology.org/	[31]
University of Michigan	http://dot.ped.med.umich.edu:2000/ourimage/pub/shared/JMR_pub_affyannot.html	[11]
NetAffx	http://www.affymetrix.com/analysis/index.affx	[12]



DAVID (Database for Annotation, Visualization, and Integrated Discovery)

Genome Biology

Software

Open Access

DAVID: Database for Annotation, Visualization, and Integrated Discovery

Glynn Dennis Jr*, Brad T Sherman*, Douglas A Hosack*, Jun Yang*,
Wei Gao*, H Clifford Lane[†] and Richard A Lempicki*

14 August, 2003 4,145 citations

Nucleic Acids Research

DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists

Da Wei Huang¹, Brad T. Sherman¹, Qina Tan¹, Joseph Kir¹, David Liu²,
David Bryant², Yongjian Guo⁵, Robert Stephens², Michael W. Baseler³,
H. Clifford Lane⁴ and Richard A. Lempicki^{1,*}

22 January, 2007 781 citations

DAVID

practices. Therefore, the DAVID Bioinformatics Tools, as web-based applications on a Tomcat web server in a Linux machine (4-CPU for 3.5 GHz speed, 8 GB memory), requires no configuration and installation in the client's computers. Java is the primary language.

Table 1. Over 22 types of gene identifiers integrated by the DAVID Gene Concept within the DAVID Knowledgebase

Gene ID Type	Total ID	Unique Cluster
AFFY_ID	2254679	845117
ENTREZ_GENE_ID	1734858	1602339
GENPEPT_ACCESSION	4065385	2511637
GENBANK_ACCESSION	16828735	2409120
GENEBANK_ID	20291282	2358084
PIR_ACCESSION	282281	258079
PIR_ID	308092	266645
PIR_NREF_ID	3355759	2677404
REFSEQ_GENOMIC	1866800	1552597
REFSEQ_MRNA	645831	561447
REFSEQ_PROTEIN	1644632	1373467
REFSEQ_RNA	1364	852
UNIGENE	161138	158938
UNIPROT_ACCESSION	2864344	2097488
UNIPROT_ID	2789453	2096712
UNIREF100_ID	2552342	2088692
OFFICIAL_GENE_SYMBOL	1693151	1600906
FLYBASE_ID	27109	26642
HAMAP_ID	63925	63822
HSSP_ID	265000	258750
TIGR_ID	120117	111699
WORMBASE_ID	43675	21243
RGD_ID	25230	25060
NOT SURE	ALL IDs	

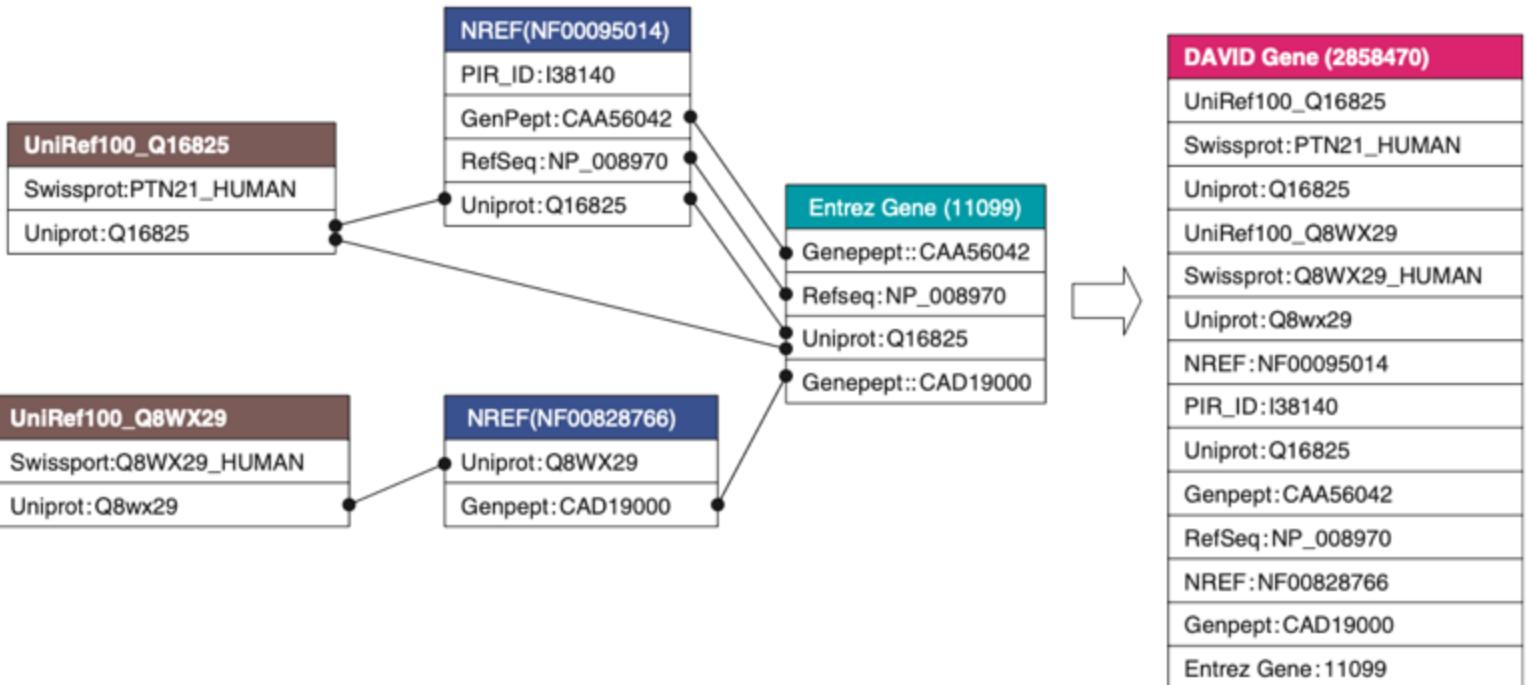


Table 2. The wide-range collection of heterogeneous functional annotations in the DAVID Knowledgebase

Ontology (>40 million records)	Protein Domain/Family (>15 millions)	Sequence Features (>21 millions)
GO_BIOLOGICAL_PROCESS	BLOCKS_ID	ALIAS_GENE_SYMBOL
GO_MOLECULAR_FUNCTION	COG_KOG_NAME	CHROMOSOME
GO_CELLULAR_COMPONENT	INTERPRO_NAME	CYTOBAND
PANTHER_BIOLOGICAL_PROCESS	PDB_ID	GENE_NAME
PANTHER_MOLECULAR_FUNCTION	PFAM_NAME	GENE_SYMBOL
COG_KOG_ONTOLOGY	PIR_ALN	HOMOLOGOUS_GENE
P-P Interaction (>4 millions)	PIR_HOMOLOGY_DOMAIN	ENTREZ_GENE_SUMMARY
BIND	PIR_SUPERFAMILY_NAME	OMIM_ID
DIP	PRINTS_NAME	PIR_SUMMARY
MINT	PRODOM_NAME	PROTEIN_MW
NCICB_CAPATHWAY	PROSITE_NAME	REFSEQ_PRODUCT
TRANSFAC_ID	SCOP_ID	SEQUENCE_LENGTH
HIV_INTERACTION	SMART_NAME	SP_COMMENT
HIV_INTERACTION_CATEGORY	TIGRFAMS_NAME	Functional Category (>6.9 millions)
HPRD_INTERACTION	PANTHER_SUBFAMILY	PIR_SEQ_FEATURE
REACTOME_INTERACTION	PANTHER_FAMILY	SP_COMMENT_TYPE
Disease Association (~9,000)	Pathways (>50 000)	SP_PIR_KEYWORDS
GENETIC_ASSOCIATION_DB	BioCarta	UP_SEQ_FEATURE
OMIM_DISEASE	KEGG_PATHWAY	Gene Tissue Expression (>1.0 million)
Literature (>2.8 millions)	PANTHER_PATHWAY	GNF Microarray
GENERIF_SUMMARY	PID	UNIGENE EST
PUBMED_ID	BBID	CGAP SAGE
HIV_INTERACTION_PUBMED_ID	KEGG_REACTION	CGAP EST

DAVID

A Hypothetical Example:

In the human genome background (*30,000 genes total; Population Total (PT)*), 40 genes are involved in the p53 signaling pathway (*Population Hits (PH)*). A given gene list has found that three genes (*List Hits (LH)*) out of 300 total genes in the list (*List Total (LT)*) belong to the p53 signaling pathway. Then we ask the question if 3/300 is more than random chance compared to the human background of 40/30000.

A 2x2 contingency table is built based on the above numbers:

List Hits (LH) = 3

List Total (LT) = 300

Population Hits (PH) = 40

Population Total (PT) = 30,000

		User Genes	Genome	
		LH	PH-LH	PH
In Pathway				
	Not In Pathway	LT-LH	PT-LT-(PH-LH)	PT-PH
		LT	PT-LT	PT

		User Genes	Genome	
		3	37	40
In Pathway				
	Not In Pathway	297	29663	29960
		300	29700	30000

Exact P-Value = 0.007. Since P-Value < 0.05, this user's gene list is specifically associated (enriched) in the p53 signaling pathway by more than random chance.

DAVID

What about the EASE Score?

The EASE Score is more conservative by subtracting one gene from the List Hits (LH) as seen below. If LH=1 (only one gene in the user's list annotated to the term), EASE Score is automatically set to 1.

	User Genes	Genome	
In Pathway	LH-1	PH-LH+1	PH
Not In Pathway	LT-LH	PT-LT-(PH-LH)	PT-PH
	LT-1	PT-LT+1	PT

	User Genes	Genome	
In Pathway	3-1	37+1	40
Not In Pathway	297	29663	29960
	300-1	29700+1	30000

For our hypothetical example involving the p53 signaling pathway, the EASE Score is more conservative with a P-value = 0.06 (using 3-1 instead of 3). Since the P-Value > 0.05, this user's gene list is not considered specifically associated (enriched) in the p53 signaling pathway any more than by random chance.

DAVID

DAVID: Database for Annotation, Visualization, and Integrated Discovery (Laboratory of Immunoprotection - Windows Internet Explorer)

http://david.abcc.ncifcrf.gov/term2term.jsp?annot=GOTERM_CC_3,GOTERM_MF_3,INTERPRO_NAME,PFAM_NAME,SMART_NAME,BIOCARTA,KEGG_PATHWAY¤tList=0

File Edit View Favorites Tools Help

Home Feeds (3) Print Page Tools Help Research

DAVID Bioinformatics Resources 2007
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Functional Annotation Clustering

Help and Manual

Current Gene List: demolist2
394 DAVID IDs

Options Classification Stringency Medium

Rerun using options Create Sublist

Download File

hyperlinks

Annotation Cluster	Enrichment Score	Category	Count	P_Value
Annotation Cluster 1	2.68	G		
GOTERM_MF_3	cytokine activity	RT	16	2.4E-4
KEGG_PATHWAY	CYTOKINE-CYTOKINE RECEPTOR INTERACTION	RT	21	3.2E-4
GOTERM_MF_3	receptor binding	RT	25	1.2E-2
GOTERM_MF_3	growth factor activity	RT	10	2.0E-2
Annotation Cluster 2	2.5	G		
GOTERM_MF_3	cytokine activity	RT	16	2.4E-4
PFAM_NAME	IL8	RT	6	1.2E-3
INTERPRO_NAME	Small chemokine, interleukin-8-like	RT	6	1.3E-3
SMART_NAME	SCY	RT	6	2.7E-3
GOTERM_MF_3	G-protein-coupled receptor binding	RT	6	7.5E-3
INTERPRO_NAME	Small chemokine, C-C	RT	4	1.3E-2
INTERPRO_NAME	Small chemokine, C-X-C/Interleukin 8	RT	3	3.3E-2
Annotation Cluster 3	2.08	G		
GOTERM_CC_3	cytoplasmic vesicle	RT	11	2.3E-3
GOTERM_CC_3	membrane-bound vesicle	RT	11	2.5E-3
GOTERM_CC_3	coated membrane	RT	4	1.0E-1
Annotation Cluster 4	1.98	G		
INTERPRO_NAME	Basic-leucine zipper (bZIP) transcription factor	RT	6	4.0E-3
SMART_NAME	BRLZ	RT	6	8.5E-3

DAVID

DAVID Functional
Annotation Clustering

Measure relationships
between annotated terms
based on the association
of their genes

Group similar and
redundant annotations
from the same or different
resources

DAVID: Database for Annotation, Visualization, and Integrated Discovery (Laboratory of Immunoprotection - Windows Internet Explorer)

http://david.ncifcrf.gov/term2term.jsp?annot=GOTERM_CC_3,GOTERM_MF_3,INTERPRO_NAME,PFAM_NAME,SMART_NAME,BIOCARTA,KEGG_PATHWAY¤tList=0

File Edit View Favorites Tools Help

Home Feeds (3) Print Page Tools Help Research

DAVID Bioinformatics Resources 2007
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Functional Annotation Clustering

Help and Manual

Current Gene List: demolist2
394 DAVID IDs

Options Classification Stringency Medium
Rerun using options Create Sublist

Download File

hyperlinks

Annotation Cluster	Enrichment Score	Term	Count	P_Value
Annotation Cluster 1	2.68	GOTERM_MF_3 cytokine activity RT	16	2.4E-4
		KEGG_PATHWAY CYTOKINE-CYTOKINE RECEPTOR INTERACTION RT	21	3.2E-4
		GOTERM_MF_3 receptor binding RT	25	1.2E-2
		GOTERM_MF_3 growth factor activity RT	10	2.0E-2
Annotation Cluster 2	2.5	GOTERM_MF_3 cytokine activity RT	16	2.4E-4
		PFAM_NAME IL6 RT	6	1.2E-3
		INTERPRO_NAME Small chemokine, interleukin-8-like RT	6	1.3E-3
		SMART_NAME SCY RT	6	2.7E-3
		GOTERM_MF_3 G-protein-coupled receptor binding RT	6	7.5E-3
		INTERPRO_NAME Small chemokine, C-C RT	4	1.3E-2
		INTERPRO_NAME Small chemokine, C-X-C/Interleukin 8 RT	3	3.3E-2
Annotation Cluster 3	2.08	GOTERM_CC_3 cytoplasmic vesicle RT	11	2.3E-3
		GOTERM_CC_3 membrane-bound vesicle RT	11	2.5E-3
		GOTERM_CC_3 coated membrane RT	4	1.0E-1
Annotation Cluster 4	1.98	INTERPRO_NAME Basic-leucine zipper (bZIP) transcription factor RT	6	4.0E-3
		SMART_NAME BRLZ RT	6	8.5E-3

DAVID (Database for Annotation, Visualization, and Integrated Discovery)

Genome Biology

Software

Open Access

DAVID: Database for Annotation, Visualization, and Integrated Discovery

Glynn Dennis Jr*, Brad T Sherman*, Douglas A Hosack*, Jun Yang*, Wei Gao*, H Clifford Lane[†] and Richard A Lempicki*

14 August, 2003 4,145 citations

Nucleic Acids Research

DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists

Da Wei Huang¹, Brad T. Sherman¹, Qina Tan¹, Joseph Kir¹, David Liu², David Bryant², Yongjian Guo⁵, Robert Stephens², Michael W. Baseler³, H. Clifford Lane⁴ and Richard A. Lempicki^{1,*}

22 January, 2007 781 citations

Genome Biology

Software

Open Access

The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists

Da Wei Huang^{**}, Brad T Sherman^{**}, Qina Tan*, Jack R Collins[†], W Gregory Alvord[‡], Jean Roayaei[‡], Robert Stephens[†], Michael W Baseler[§], H Clifford Lane[¶] and Richard A Lempicki*

4 September, 2007 887 citations

DAVID

Measure functional relations of gene-pairs based on their annotations

(a)

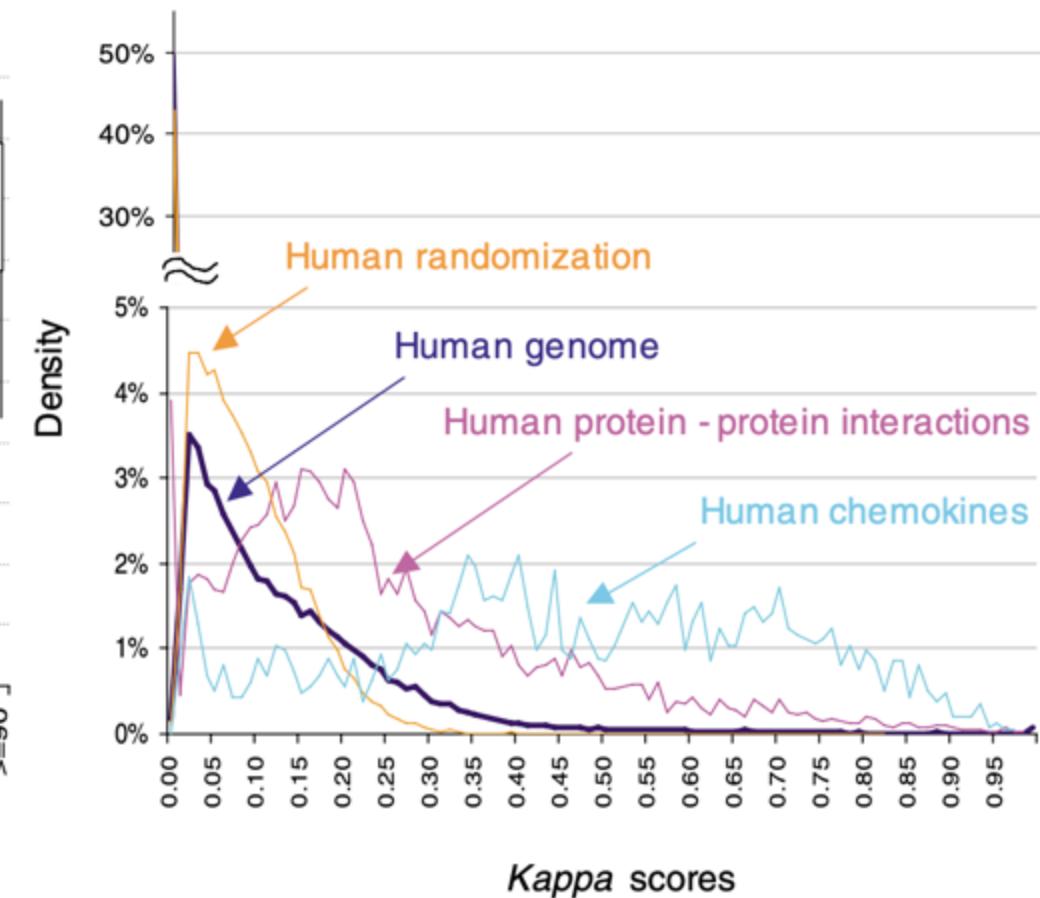
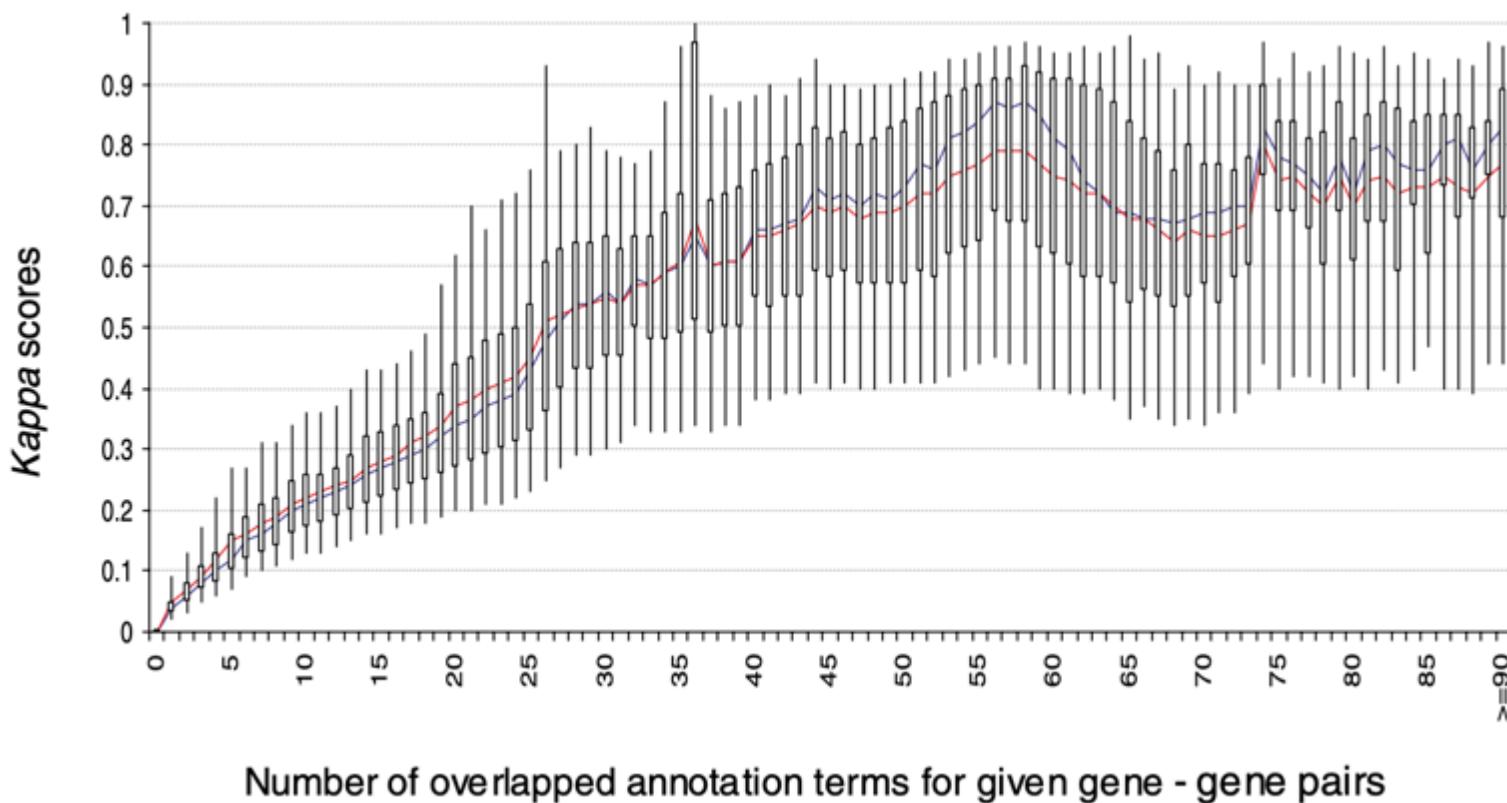
	Cell death	Apoptosis	Ph domain	Sh2 domain	Apoptosis pathway	Membrane
Gene a	1	1	0	0	1	0
Gene b	1	1	0	1	1	0
Gene c	1	0	0	1	1	1
Gene d	1	1	0	0	1	1
Gene e	0	1	1	1	1	1
Gene f	0	0	1	1	0	1
Gene g	0	0	1	1	0	1

(b)

		Gene a		Row total
		1	0	
Gene b	1	3 ($C_{1,1}$)	1 ($C_{0,1}$)	4 ($C_{1,\cdot}$)
	0	0 ($C_{0,1}$)	2 ($C_{0,0}$)	2 ($C_{0,\cdot}$)
Column total		3 ($C_{\cdot,1}$)	3 ($C_{\cdot,0}$)	6 (T_{ab})
Observed co-occurrence				
$O_{ab} = \frac{C_{1,1} + C_{0,0}}{T_{ab}} = \frac{3 + 2}{6} = 0.83$				
Chance of co-occurrence				
$A_{ab} = \frac{C_{\cdot,1} \cdot C_{1,\cdot} + C_{\cdot,0} \cdot C_{0,\cdot}}{T_{ab} \cdot T_{ab}} = \frac{3 \cdot 4 + 3 \cdot 2}{6 \cdot 6} = 0.5$				
$K_{ab} = \frac{O_{ab} - A_{ab}}{1 - A_{ab}} = \frac{0.83 - 0.5}{1 - 0.5} = 0.66$				

Degree of annotation co-occurrence between genes a and b

DAVID



DAVID

Agglomeration method to
classify a gene list into
functionally related
groups based on the
functional similarity scores

DAVID

Agglomeration method to
classify a gene list into
functionally related
groups based on the
functional similarity scores



K-means, FANNY and SOM

DAVID

Agglomeration method to
classify a gene list into
functionally related
groups based on the
functional similarity scores



~~K means, FANNY and SOM~~



Heuristic fuzzy multiple-
linkage partitioning
agglomeration approach

DAVID

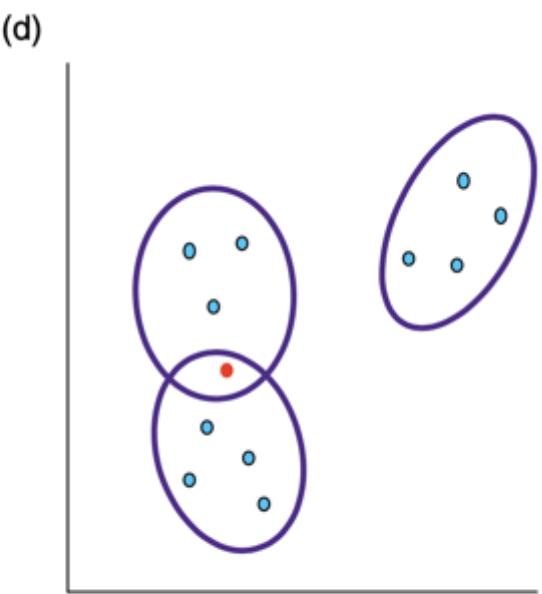
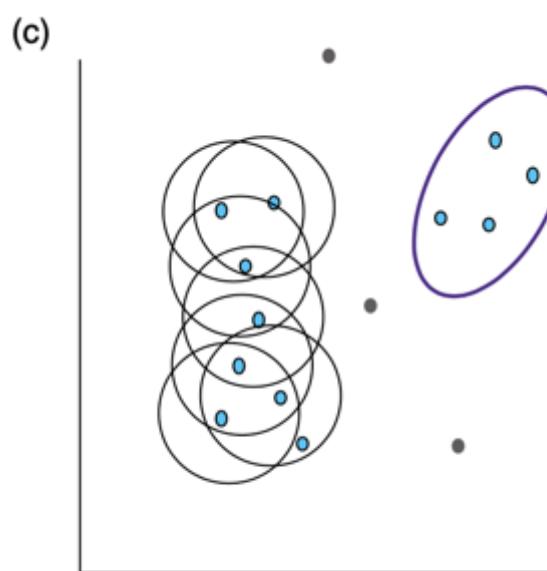
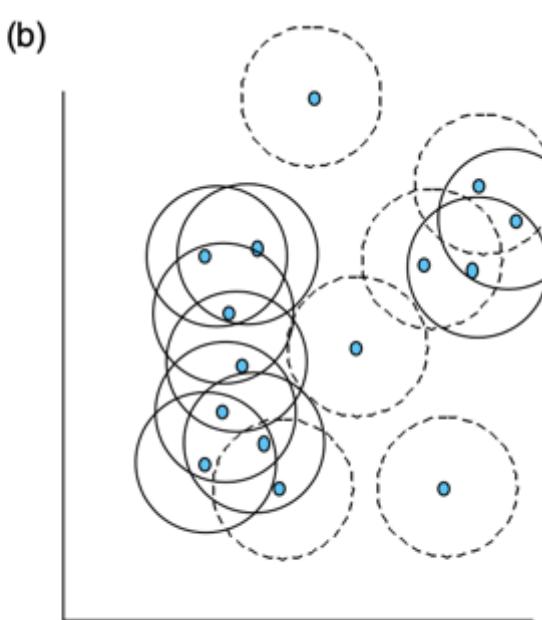
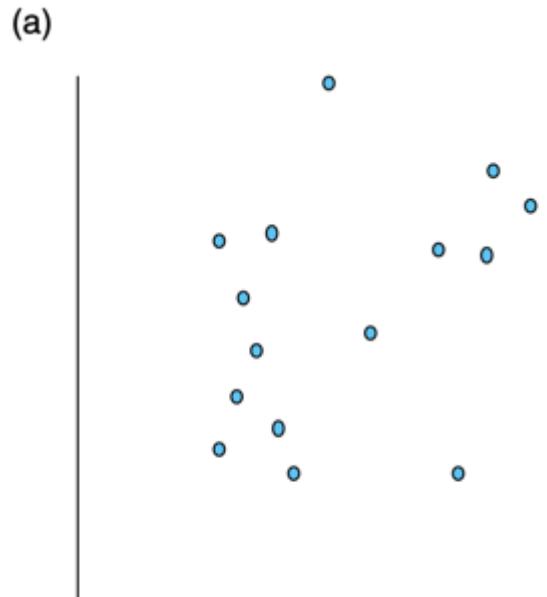
Agglomeration method to classify a gene list into functionally related groups based on the functional similarity scores



~~K means, FANNY and SOM~~



Heuristic fuzzy multiple-linkage partitioning agglomeration approach



DAVID

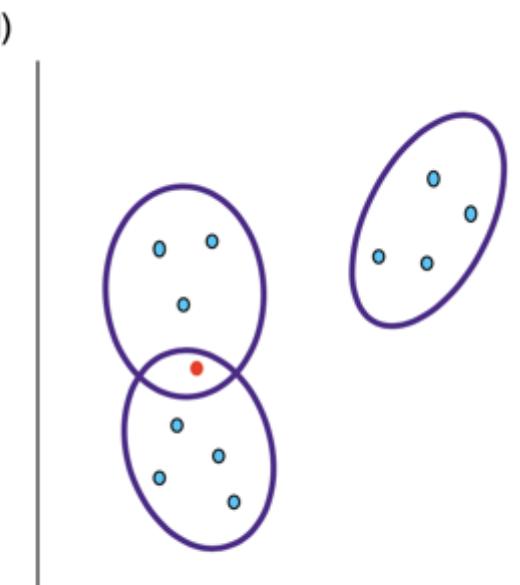
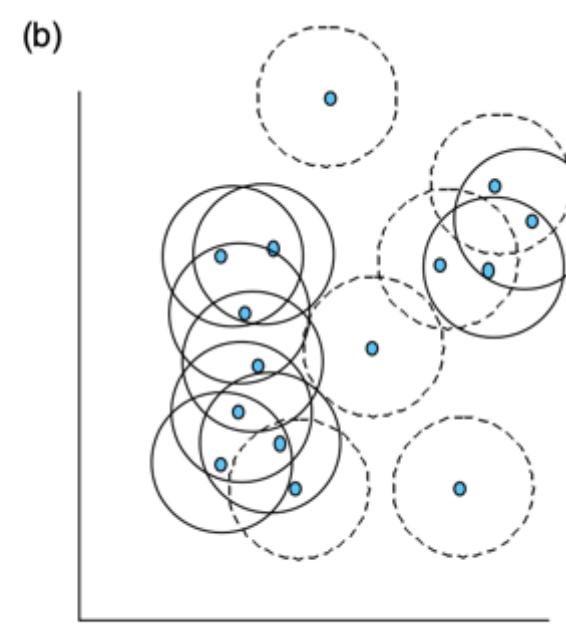
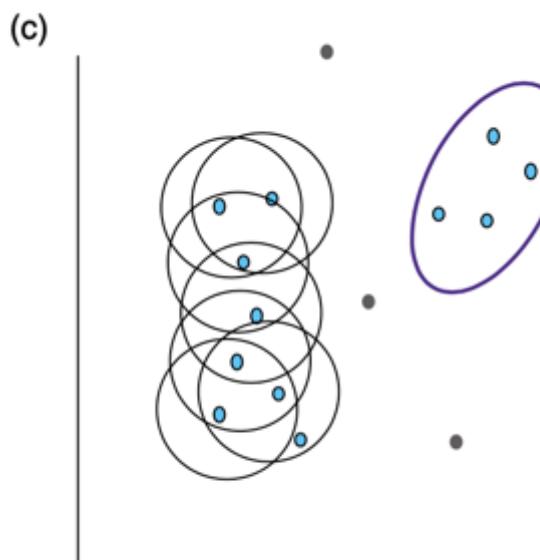
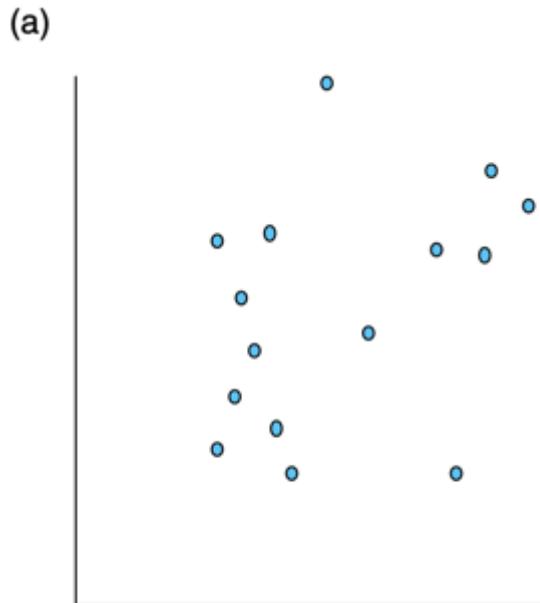
Agglomeration method to classify a gene list into functionally related groups based on the functional similarity scores



~~K means, FANNY and SOM~~



Heuristic fuzzy multiple-linkage partitioning agglomeration approach



Evaluate the quality of the agglomeration algorithm



Randomly select genes that belong to the same protein family (kinases, phosphatases...)



Classify the genes

Genes not classified into the group they belong to

1-2%

Genes classified into groups they don't belong to

2-5%

DAVID

Which functional gene groups are more significant for the experiment?
(order the groups by their relative importance)



A gene group is more important if the majority of the genes are associated with highly enriched terms



Geometric mean of the EASE scores associated with the terms that belong to the gene group
(both significant and non-significant)



Minus log transformation

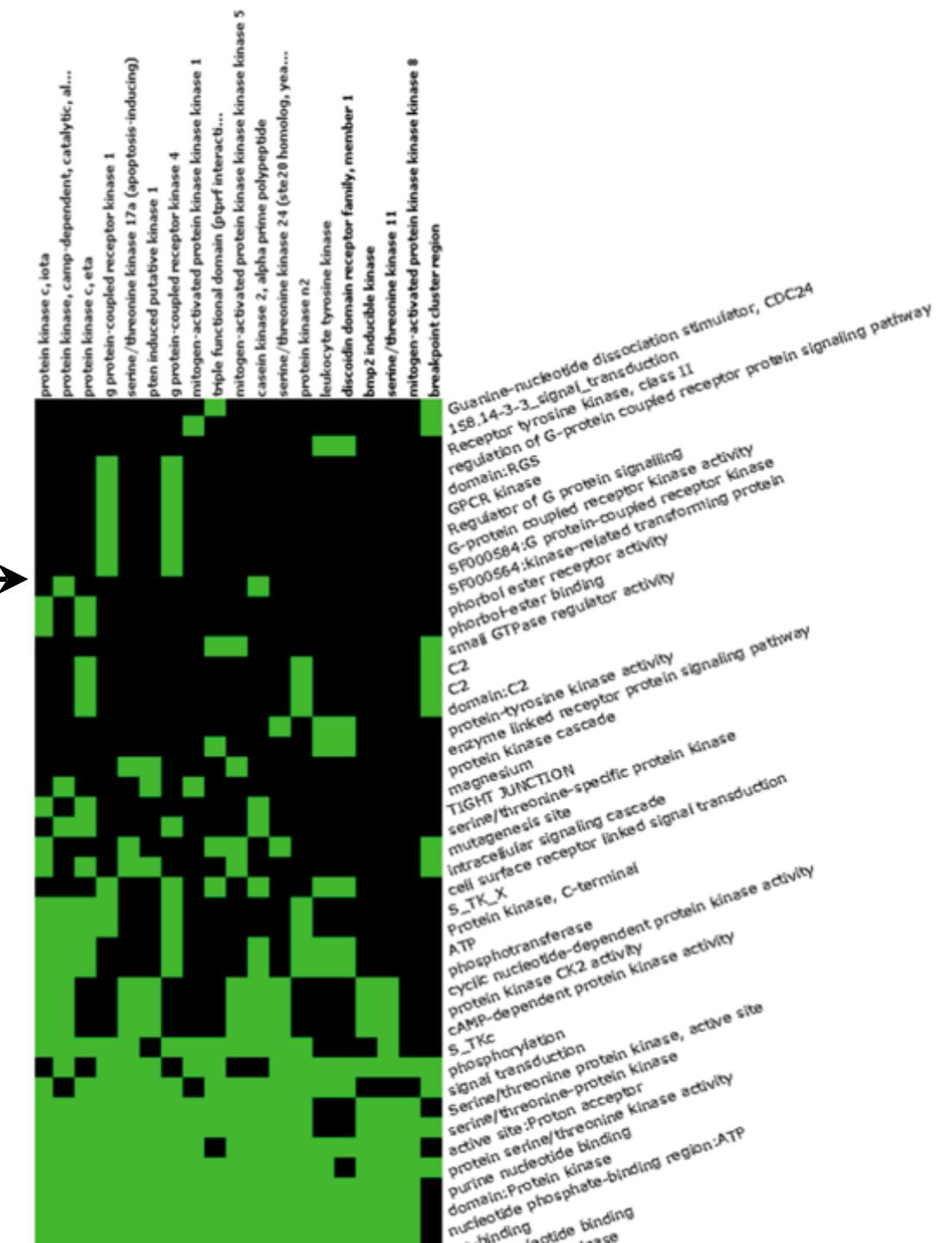


Group enrichment score

	User Genes	Genome	
In Pathway	LH-1	PH-LH+1	PH
Not In Pathway	LT-LH	PT-LT-(PH-LH)	PT-PH
	LT-1	PT-LT+1	PT
	User Genes	Genome	
In Pathway	3-1	37+1	40
Not In Pathway	297	29663	29960
	300-1	29700+1	30000

DAVID

Gene Group 1		Enrichment Score: 3.78	RG	T	
1	34375_at, 875_g_at	chemokine (c-c motif) ligand 2			
2	40385_at	chemokine (c-c motif) ligand 20			
3	36103_at	chemokine (c-c motif) ligand 3			
4	36674_at	chemokine (c-c motif) ligand 4			
5	408_at	chemokine (c-x-c motif) ligand 1 (melanoma growth stimulating activity, alpha)			
Gene Group 2		Enrichment Score: 2.68	RG	T	
1	34475_at	g_protein-coupled receptor kinase 4			
2	568_at	protein kinase, camp-dependent, catalytic, alpha			
3	1267_at	protein kinase c, eta			
4	34226_at	mitogen-activated protein kinase kinase kinase 5			
5	37629_at	casein kinase 2, alpha prime polypeptide			
Gene Group 3		Enrichment Score: 1.34	RG	T	
1	39748_at	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1			
2	32186_at	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5			
3	33143_s_at	solute carrier family 16 (monocarboxylic acid transporters), member 3			
4	39841_at	solute carrier family 16 (monocarboxylic acid transporters), member 6			



DAVID

DAVID Functional
Annotation Clustering

Measure relationships
between annotated terms
based on the association
of their genes

Group similar and
redundant annotations
from the same or different
resources

DAVID: Database for Annotation, Visualization, and Integrated Discovery (Laboratory of Immunoprotection - Windows Internet Explorer)

http://david.ncifcrf.gov/term2term.jsp?annot=GOTERM_CC_3,GOTERM_MF_3,INTERPRO_NAME,PFAM_NAME,SMART_NAME,BIOCARTA,KEGG_PATHWAY¤tList=0

File Edit View Favorites Tools Help

Home Feeds (3) Print Page Tools Help Research

DAVID Bioinformatics Resources 2007
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Functional Annotation Clustering

Help and Manual

Current Gene List: demolist2
394 DAVID IDs

Options Classification Stringency Medium
Rerun using options Create Sublist

Download File

hyperlinks

Annotation Cluster	Enrichment Score	Term	Count	P_Value
Annotation Cluster 1	2.68	GOTERM_MF_3 cytokine activity RT	16	2.4E-4
		GOTERM_MF_3 CYTOKINE-CYTOKINE RECEPTOR INTERACTION RT	21	3.2E-4
		GOTERM_MF_3 receptor binding RT	25	1.2E-2
		GOTERM_MF_3 growth factor activity RT	10	2.0E-2
Annotation Cluster 2	2.5	GOTERM_MF_3 cytokine activity RT	16	2.4E-4
		PFAM_NAME IL6 RT	6	1.2E-3
		INTERPRO_NAME Small chemokine, interleukin-8-like RT	6	1.3E-3
		SMART_NAME SCY RT	6	2.7E-3
		GOTERM_MF_3 G-protein-coupled receptor binding RT	6	7.5E-3
		INTERPRO_NAME Small chemokine, C-C RT	4	1.3E-2
		INTERPRO_NAME Small chemokine, C-X-C/Interleukin 8 RT	3	3.3E-2
Annotation Cluster 3	2.08	GOTERM_CC_3 cytoplasmic vesicle RT	11	2.3E-3
		GOTERM_CC_3 membrane-bound vesicle RT	11	2.5E-3
		GOTERM_CC_3 coated membrane RT	4	1.0E-1
Annotation Cluster 4	1.98	INTERPRO_NAME Basic-leucine zipper (bZIP) transcription factor RT	6	4.0E-3
		SMART_NAME BRLZ RT	6	8.5E-3

DAVID



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](#) [About DAVID](#) [About LHRI](#)

Overview

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

Hot Links

[Two postdoctoral fellow positions for HIV basic science research available](#)

The Laboratory of Human Retrovirology and Immunoinformatics (LHRI), Applied/ Developmental Directorate, has investigated mechanisms of multiple-class drug resistant (MDR) in HIV infection and regulation of autoprocessing/budding in HIV replication. We have two postdoctoral fellow positions available to perform Molecular Virology and HIV Virology research in our [Basic Research Section](#).

[Call for papers](#)

Submit papers for a Special Issue:"DNA or RNA-Mediated Innate Immune Response" of the International Journal of Molecular Sciences

[DAVID Forum](#)

Forum for DAVID users to ask questions, suggest new functions and help other users by answering their questions.

[FAQ](#)

Frequently Asked Questions

[LHRI Publications](#)

Publications of the Laboratory of Human Retrovirology and Immunoinformatics, Frederick National Laboratory for Cancer Research

[DAVID Publications](#)

Publications about DAVID

DAVID



Analysis Wizard

DAVID Bioinformatics Resources 6.8, NIAID/NIH

[Home](#) [Start Analysis](#) [Shortcut to DAVID Tools](#) [Technical Center](#) [Downloads & APIs](#) [Term of Service](#) [About DAVID](#) [About LHRI](#)

[Upload](#) [List](#) [Background](#)

Upload Gene List

[Demolist 1](#) [Demolist 2](#)

[Upload Help](#)

Step 1: Enter Gene List

A: Paste a list

```
ENSG00000145216
ENSG00000183426
ENSG00000134759
ENSG00000132842
```

[Clear](#)

Or
B: Choose From a File
 [no file selected](#)
 Multi-List File [?](#)

Step 2: Select Identifier

ENSEMBL_GENE_ID [?](#)

An example:

Copy/paste IDs to "box A" -> Select Identifier as "Affy_ID" -> List Type as "Gene List" -> Click "Submit"

```
1007_s_at
1053_at
117_at
121_at
1255_g_at
1294_at
1316_at
1320_at
1405_i_at
1431_at
1438_at
1487_at
1494_f_at
1598_g_at
```

Step 3: List Type

Gene List [?](#)

Background [?](#)

Step 4: Submit List

[Submit List](#)

Analysis Wizard

Tell us how you like the tool
[Contact us for questions](#)

← Step 1. Submit your gene list through left panel.

DAVID



Analysis Wizard

DAVID Bioinformatics Resources 6.8, NIAID/NIH

[Home](#) [Start Analysis](#) [Shortcut to DAVID Tools](#) [Technical Center](#) [Downloads & APIs](#) [Term of Service](#) [About DAVID](#) [About LHRI](#)

[Upload](#) [List](#) [Background](#)

Gene List Manager

Select to limit annotations by one or more species [Help](#)

- Use All Species -
Homo sapiens(4050)
Unknown(146)

[Select Species](#)

[List Manager](#) [Help](#)

List_1

Select List to:

[Use](#) [Rename](#)

[Remove](#) [Combine](#)

[Show Gene List](#)

[View Unmapped Ids](#)

Analysis Wizard

Tell us how you like the tool
[Contact us for questions](#)

Step 1. Successfully submitted gene list

Current Gene List: List_1

Current Background: Homo sapiens

Step 2. Analyze above gene list with one of DAVID tools



[Which DAVID tools to use?](#)

Functional Annotation Tool

- Functional Annotation Clustering
- Functional Annotation Chart
- Functional Annotation Table

Gene Functional Classification Tool

Gene ID Conversion Tool

Gene Name Batch Viewer

DAVID

Analysis Wizard
DAVID Bioinformatics Resources 6.8, NIAID/NIH

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

Upload List Background

Upload Gene List

Demolist 1 Demolist 2
Upload Help

Step 1: Enter Gene List

A: Paste a list

```
ENSG00000170323
ENSG00000173432
ENSG00000166819
ENSG00000162896
```

Clear

Or

B: Choose From a File

Choose File no file selected

Multi-List File ?

Step 2: Select Identifier

AFFYMETRIX_3PRIME_IVT_ID
AFFYMETRIX_EXON_ID
AGILENT_CHIP_ID
AGILENT_ID
AGILENT_OLIGO_ID
APHIDBASE_ID
BEEBASE_ID
BEETLEBASE_ID
BGD_ID
CGNC_ID
CRYPTODB_ID
DICTYBASE_ID
ENSEMBL_GENE_ID
ENSEMBL_TRANSCRIPT_ID
ENTREZ_GENE_ID

Analysis Wizard

Tell us how you like the tool
Contact us for questions

Step 1. Successfully submitted gene list
Current Gene List: BreastC_up
Current Background: Homo sapiens

Step 2. Analyze above gene list with one of DAVID tools

Which DAVID tools to use?

↓

- ⊕ Functional Annotation Tool
 - Functional Annotation Clustering
 - Functional Annotation Chart
 - Functional Annotation Table
- ⊕ Gene Functional Classification Tool
- ⊕ Gene ID Conversion Tool
- ⊕ Gene Name Batch Viewer

DAVID

DAVID Bioinformatics Database

Analysis Wizard
DAVID Bioinformatics Resources 6.8, NIAID/NIH

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

Upload List Background

Upload Gene List

Demolist 1 Demolist 2
Upload Help

Step 1: Enter Gene List

A: Paste a list

ENSG00000170323
ENSG00000173432
ENSG00000166819
ENSG00000162896

Clear

Or

B: Choose From a File
Choose File no file selected
Multi-List File ?

Step 2: Select Identifier

ENSEMBL_GENE_ID

Step 3: List Type

Gene List
Background

Step 4: Submit List

Submit List

Analysis Wizard

Tell us how you like the tool
[Contact us for questions](#)

Step 1. Successfully submitted gene list
Current Gene List: BreastC_up
Current Background: Homo sapiens

Step 2. Analyze above gene list with one of DAVID tools

Which DAVID tools to use?

↓

- ⊕ Functional Annotation Tool
 - Functional Annotation Clustering
 - Functional Annotation Chart
 - Functional Annotation Table
- ⊕ Gene Functional Classification Tool
- ⊕ Gene ID Conversion Tool
- ⊕ Gene Name Batch Viewer

DAVID



Functional Annotation Tool

DAVID Bioinformatics Resources 6.8, NIAID/NIH

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

Upload List Background

Gene List Manager

Select to limit annotations by one or more species [Help](#)

- Use All Species -
 - Homo sapiens(4050)
 - Unknown(146)

[Select Species](#)

[List Manager](#) [Help](#)

BreastC_up

Select List to:

[Use](#) [Rename](#)

[Remove](#) [Combine](#)

[Show Gene List](#)

[View Unmapped Ids](#)

Annotation Summary Results

[Help and Tool Manual](#)

Current Gene List: BreastC_up

4047 DAVID IDs

Current Background: Background_2

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](#)

Upload List Background

Gene List Manager

Select to limit annotations by one or more species [Help](#)

- Use All Species -
Homo sapiens(4050)
Unknown(146)

Select Species

List Manager [Help](#)
BreastC_up

Select List to:

Use Rename
Remove Combine
Show Gene List

View Unmapped Ids

Annotation Summary Results

Current Gene List: BreastC_up
Current Background: Background_2

4047 DAVID IDs
 Check Defaults

Disease (1 selected)
 Functional_Categories (3 selected)
 Gene_Ontology (3 selected)

<input type="checkbox"/> GOTERM_BP_1	81.2%	3288	Chart	
<input type="checkbox"/> GOTERM_BP_2	81.1%	3284	Chart	
<input type="checkbox"/> GOTERM_BP_3	80.7%	3265	Chart	
<input type="checkbox"/> GOTERM_BP_4	79.7%	3227	Chart	
<input type="checkbox"/> GOTERM_BP_5	78.9%	3193	Chart	
<input type="checkbox"/> GOTERM_BP_ALL	81.2%	3288	Chart	
<input checked="" type="checkbox"/> GOTERM_BP_DIRECT	81.2%	3288	Chart	
<input type="checkbox"/> GOTERM_BP_FAT	80.8%	3270	Chart	
<input type="checkbox"/> GOTERM_CC_1	86.3%	3493	Chart	
<input type="checkbox"/> GOTERM_CC_2	85.9%	3475	Chart	
<input type="checkbox"/> GOTERM_CC_3	85.8%	3474	Chart	
<input type="checkbox"/> GOTERM_CC_4	85.0%	3438	Chart	
<input type="checkbox"/> GOTERM_CC_5	80.9%	3273	Chart	
<input type="checkbox"/> GOTERM_CC_ALL	86.3%	3493	Chart	
<input checked="" type="checkbox"/> GOTERM_CC_DIRECT	86.3%	3493	Chart	
<input type="checkbox"/> GOTERM_CC_FAT	73.2%	2963	Chart	
<input type="checkbox"/> GOTERM_MF_1	82.5%	3338	Chart	
<input type="checkbox"/> GOTERM_MF_2	82.4%	3333	Chart	
<input type="checkbox"/> GOTERM_MF_3	71.8%	2904	Chart	
<input type="checkbox"/> GOTERM_MF_4	68.3%	2765	Chart	
<input type="checkbox"/> GOTERM_MF_5	55.5%	2247	Chart	
<input type="checkbox"/> GOTERM_MF_ALL	82.5%	3338	Chart	
<input checked="" type="checkbox"/> GOTERM_MF_DIRECT	82.5%	3338	Chart	
<input type="checkbox"/> GOTERM_MF_FAT	73.2%	2964	Chart	

General_Annotations (0 selected)
 Literature (0 selected)
 Main_Accessions (0 selected)
 Pathways (3 selected)

<input checked="" type="checkbox"/> BBID	1.6%	63	Chart	
<input checked="" type="checkbox"/> BIOCARTA	7.2%	291	Chart	
<input type="checkbox"/> EC_NUMBER	22.9%	926	Chart	
<input checked="" type="checkbox"/> KEGG_PATHWAY	31.4%	1271	Chart	
<input type="checkbox"/> REACTOME_PATHWAY	46.8%	1892	Chart	

[Help and Tool Manual](#)

Functional Annotation Chart

[Help and Manual](#)
Current Gene List: BreastC_up

Current Background: Background_2

4047 DAVID IDs
 [Options](#)
[Rerun Using Options](#) [Create Sublist](#)
[Download File](#)
24 chart records

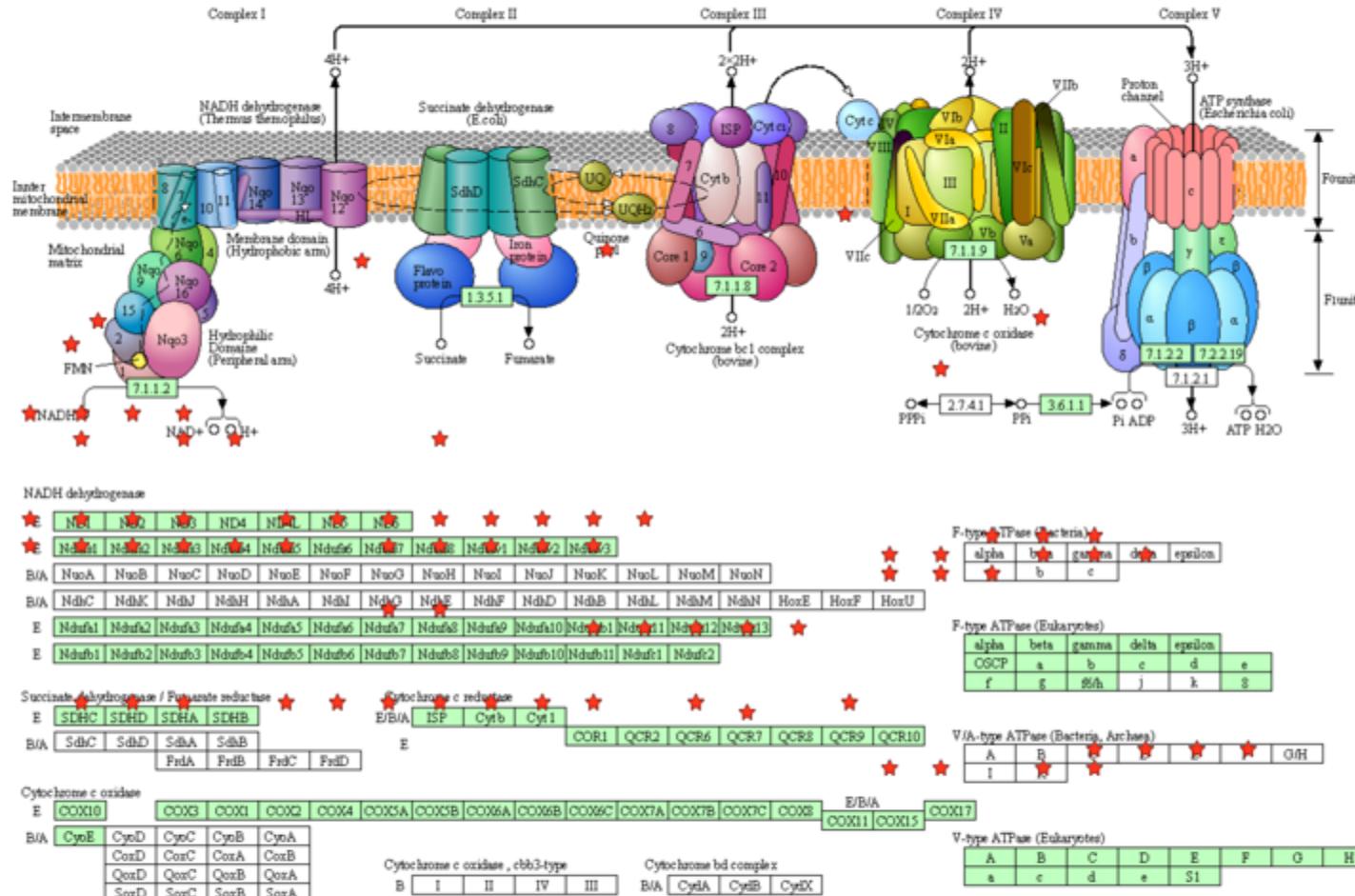
Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	KEGG_PATHWAY	Oxidative phosphorylation	RT	77	1.9	6.7E-17	1.9E-14	
	KEGG_PATHWAY	Parkinson's disease	RT	70	1.7	1.5E-11	2.2E-9	
	KEGG_PATHWAY	Huntington's disease	RT	86	2.1	8.8E-11	8.4E-9	
	KEGG_PATHWAY	Alzheimer's disease	RT	77	1.9	2.3E-10	1.6E-8	
	KEGG_PATHWAY	DNA replication	RT	27	0.7	1.7E-8	9.5E-7	
	KEGG_PATHWAY	Fanconi anemia pathway	RT	32	0.8	9.6E-8	4.5E-6	
	KEGG_PATHWAY	Base excision repair	RT	23	0.6	1.9E-6	7.8E-5	
	KEGG_PATHWAY	Spliceosome	RT	62	1.5	2.8E-6	1.0E-4	
	KEGG_PATHWAY	Pyrimidine metabolism	RT	45	1.1	4.7E-6	1.5E-4	
	KEGG_PATHWAY	RNA transport	RT	65	1.6	2.4E-5	6.9E-4	
	KEGG_PATHWAY	Ribosome biogenesis in eukaryotes	RT	35	0.9	1.2E-4	3.1E-3	
	KEGG_PATHWAY	Metabolic pathways	RT	302	7.5	1.6E-4	3.7E-3	
	KEGG_PATHWAY	Cell cycle	RT	47	1.2	2.1E-4	4.6E-3	
	KEGG_PATHWAY	Non-alcoholic fatty liver disease (NAFLD)	RT	57	1.4	2.6E-4	5.2E-3	
	KEGG_PATHWAY	Cardiac muscle contraction	RT	22	0.5	1.8E-3	3.3E-2	
	KEGG_PATHWAY	Mismatch repair	RT	14	0.3	1.9E-3	3.5E-2	
	KEGG_PATHWAY	Ribosome	RT	52	1.3	3.8E-3	6.3E-2	
	KEGG_PATHWAY	Systemic lupus erythematosus	RT	18	0.4	4.0E-3	6.4E-2	
	KEGG_PATHWAY	Purine metabolism	RT	55	1.4	1.8E-2	2.6E-1	
	KEGG_PATHWAY	Protein export	RT	13	0.3	1.8E-2	2.6E-1	
	KEGG_PATHWAY	Homologous recombination	RT	12	0.3	2.2E-2	3.0E-1	
	KEGG_PATHWAY	Drug metabolism - other enzymes	RT	9	0.2	6.7E-2	8.6E-1	
	KEGG_PATHWAY	Amino sugar and nucleotide sugar metabolism	RT	18	0.4	9.9E-2	1.0E0	
	KEGG_PATHWAY	Nucleotide excision repair	RT	18	0.4	9.9E-2	1.0E0	

3358 gene(s) from your list are not in the output.

DAVID

Pathway:Oxidative phosphorylation Pathway information generated by KEGG. Stop Blinking

OXIDATIVE PHOSPHORYLATION



00190 7/7/20

List genes are shown in red

DAVID Gene Name

ATP synthase F0 subunit 6(ATP6),
ATP synthase F0 subunit 8(ATP8),
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit(ATPSO),
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle(ATPSA1),
ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide(ATPSB),
ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit(ATPSD),
ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit(ATPSE),
ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1(ATPSC1),
ATP synthase, H+ transporting, mitochondrial Fo complex subunit B1(ATPSF1),
ATP synthase, H+ transporting, mitochondrial Fo complex subunit C1 (subunit 9)(ATPSG1).

Functional Related Terms

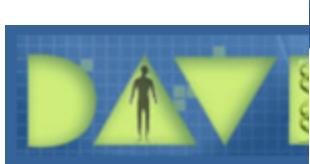
Options

Rerun using options

21943 term(s) were searched. 19 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	KEGG_PATHWAY	Oxidative phosphorylation	1.00
2	KEGG_PATHWAY	Parkinson's disease	0.84
3	KEGG_PATHWAY	Alzheimer's disease	0.75
4	KEGG_PATHWAY	Non-alcoholic fatty liver disease (NAFLD)	0.71
5	KEGG_PATHWAY	Huntington's disease	0.71
6	UP_KEYWORDS	Respiratory chain	0.68
7	UP_KEYWORDS	Electron transport	0.66
8	GOTERM_BP_DIRECT	mitochondrial electron transport, NADH to ubiquinone	0.59
9	GOTERM_CC_DIRECT	mitochondrial respiratory chain complex I	0.58
10	GOTERM_MF_DIRECT	NADH dehydrogenase (ubiquinone) activity	0.57
11	UP_KEYWORDS	Mitochondrion inner membrane	0.53
12	GOTERM_BP_DIRECT	mitochondrial respiratory chain complex I assembly	0.51
13	UP_KEYWORDS	Hydrogen ion transport	0.44
14	UP_KEYWORDS	Ubiquinone	0.41
15	GOTERM_CC_DIRECT	mitochondrial inner membrane	0.39
16	KEGG_PATHWAY	Metabolic pathways	0.38
17	KEGG_PATHWAY	Cardiac muscle contraction	0.36
18	GOTERM_BP_DIRECT	mitochondrial ATP synthesis coupled proton transport	0.32
19	GOTERM_BP_DIRECT	hydrogen ion transmembrane transport	0.31



Functional Annotation Chart

[Help and Manual](#)
 Current Gene List: BreastC_up
 Current Background: Background_2
 4047 DAVID IDs
[Options](#)

Functional Re

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

479 chart records

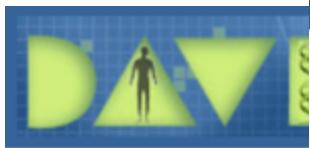
[Download File](#)[Options](#)[Rerun using options](#)

21943 term(s) wer

[Download File](#)

Similarity Score:

#			Term	RT	Genes	Count	%	P-Value	Benjamini	Kappa
1	KEGG_PATHWAY	GOTERM_CC_DIRECT	mitochondrial inner membrane	RT	210	5.2	9.6E-20	9.5E-17		1.00
2	KEGG_PATHWAY	GOTERM_BP_DIRECT	mitochondrial translational elongation	RT	65	1.6	5.0E-18	1.6E-14		0.84
3	KEGG_PATHWAY	GOTERM_BP_DIRECT	mitochondrial translational termination	RT	66	1.6	5.9E-18	1.6E-14		0.75
4	KEGG_PATHWAY	KEGG_PATHWAY	Oxidative phosphorylation	RT	77	1.9	6.7E-17	1.9E-14		0.71
5	KEGG_PATHWAY	UP_KEYWORDS	Mitochondrion	RT	414	10.2	7.0E-14	4.0E-11		0.71
6	UP_KEYWORDS	GOTERM_BP_DIRECT	DNA replication	RT	82	2.0	1.6E-13	2.8E-10		0.68
7	UP_KEYWORDS	UP_KEYWORDS	Mitochondrion inner membrane	RT	130	3.2	3.8E-12	1.1E-9		0.66
8	UP_KEYWORDS	KEGG_PATHWAY	Parkinson's disease	RT	70	1.7	1.5E-11	2.2E-9		0.59
9	GOTERM_BP_DI	UP_KEYWORDS	Respiratory chain	RT	45	1.1	1.5E-11	2.9E-9		0.58
10	GOTERM_CC_DI	KEGG_PATHWAY	Huntington's disease	RT	86	2.1	8.8E-11	8.4E-9		0.57
11	GOTERM_MF_DI	UP_KEYWORDS	DNA repair	RT	132	3.3	2.0E-10	2.8E-8		0.53
12	UP_KEYWORDS	GOTERM_BP_DIRECT	Alzheimer's disease	RT	77	1.9	2.3E-10	1.6E-8		0.51
13	GOTERM_BP_DI	UP_KEYWORDS	Mitochondrial respiratory chain complex I assembly	RT	44	1.1	2.6E-10	3.5E-7		0.44
14	UP_KEYWORDS	GOTERM_CC_DIRECT	DNA damage	RT	151	3.7	7.5E-10	8.6E-8		0.41
15	UP_KEYWORDS	GOTERM_BP_DIRECT	mitochondrial respiratory chain complex I	RT	35	0.9	1.1E-9	5.7E-7		0.39
16	GOTERM_CC_DI	KEGG_PATHWAY	mitochondrial large ribosomal subunit	RT	36	0.9	3.3E-9	1.1E-6		0.38
17	KEGG_PATHWAY	UP_KEYWORDS	Mitosis	RT	104	2.6	5.3E-9	5.0E-7		0.36
18	KEGG_PATHWAY	GOTERM_BP_DIRECT	DNA repair	RT	103	2.5	7.0E-9	7.7E-6		0.32
19	GOTERM_BP_DI	UP_KEYWORDS	mitochondrial electron transport, NADH to ubiquinone	RT	145	3.6	9.8E-9	8.0E-7		0.31
		SMART	KRAB	RT	128	3.2	1.2E-8	6.9E-6		
		GOTERM_CC_DIRECT	centriole	RT	56	1.4	1.3E-8	3.3E-6		
		UP_KEYWORDS	DNA replication	RT	50	1.2	1.4E-8	1.0E-6		
		GOTERM_BP_DIRECT	mitochondrial electron transport, NADH to ubiquinone	RT	34	0.8	1.6E-8	1.4E-5		
		KEGG_PATHWAY	DNA replication	RT	27	0.7	1.7E-8	9.5E-7		
		UP_KEYWORDS	Transit peptide	RT	206	5.1	2.1E-8	1.4E-6		


Functional Annotation Table
[Help and Manual](#)

Current Gene List: BreastC_up

Current Background: Background_2

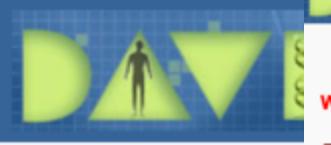
4047 DAVID IDs

Functional Re
Options
[Rerun using options](#)
21943 term(s) wer
Similarity Score:

#			Download File
1	KEGG_PATHWAY	ENSG00000089127	Download File
2	KEGG_PATHWAY	GOTERM_BP_DIRECT	glucose metabolic process, purine nucleotide biosynthetic process, immune response, response to virus, cellular response to interferon-alpha, glucose homeostasis, negative regulation of viral genome replication, protein oligomerization, defense response to virus, interferon-gamma-mediated signaling pathway, type I interferon signaling pathway, regulation of ribonuclease activity,
3	KEGG_PATHWAY	GOTERM_CC_DIRECT	extracellular region, nucleus, cytoplasm, mitochondrion, endoplasmic reticulum, cytosol,
4	KEGG_PATHWAY	GOTERM_MF_DIRECT	2'-5'-oligoadenylate synthetase activity, double-stranded RNA binding, protein binding, ATP binding, zinc ion binding, transferase activity, nucleotidyltransferase activity, metal ion binding,
5	KEGG_PATHWAY	INTERPRO	Nucleotidyl transferase domain, 2'-5'-oligoadenylate synthetase, N-terminal, 2'-5'-oligoadenylate synthetase, conserved site, 2'-5'-oligoadenylate synthetase 1, domain 2/C-terminal, 2'-5'-oligoadenylate synthetase,
6	KEGG_PATHWAY	KEGG_PATHWAY	Hepatitis C, Measles, Influenza A, Herpes simplex infection,
7	KEGG_PATHWAY	OMIM_DISEASE	Diabetes mellitus, type 1, susceptibility to, Viral infection, susceptibility to,
8	KEGG_PATHWAY	UP_KEYWORDS	3D-structure, Alternative splicing, Antiviral defense, ATP-binding, Complete proteome, Cytoplasm, Endoplasmic reticulum, Immunity, Innate immunity, Magnesium, Metal-binding, Microsome, Mitochondrion, Nucleotide-binding, Nucleotidyltransferase, Nucleus, Polymorphism, Proteomics identification, Reference proteome, RNA-binding, Secreted, Transferase,
9	KEGG_PATHWAY	UP_SEQ_FEATURE	chain:2'-5'-oligoadenylate synthetase 1, mutagenesis site, region of interest:Necessary for binding to dsRNA, sequence conflict, sequence variant, splice variant,
10	KEGG_PATHWAY	ENSG00000111335	Download File
11	KEGG_PATHWAY	GOTERM_BP_DIRECT	nucleobase-containing compound metabolic process, purine nucleotide biosynthetic process, RNA catabolic process, protein glycosylation, immune response, response to virus, protein myristylation, defense response to virus, interferon-gamma-mediated signaling pathway, type I interferon signaling pathway,
12	KEGG_PATHWAY	GOTERM_CC_DIRECT	nucleus, cytoplasm, mitochondrion, endoplasmic reticulum, cytosol, membrane, intracellular membrane-bound organelle, perinuclear region of cytoplasm,
13	KEGG_PATHWAY	GOTERM_MF_DIRECT	2'-5'-oligoadenylate synthetase activity, double-stranded RNA binding, protein binding, ATP binding, zinc ion binding, transferase activity, nucleotidyltransferase activity, metal ion binding,
14	KEGG_PATHWAY	INTERPRO	Nucleotidyl transferase domain, 2'-5'-oligoadenylate synthetase, N-terminal, 2'-5'-oligoadenylate synthetase, conserved site, 2'-5'-oligoadenylate synthetase 1, domain 2/C-terminal, 2'-5'-oligoadenylate synthetase,
15	KEGG_PATHWAY	KEGG_PATHWAY	Hepatitis C, Measles, Influenza A, Herpes simplex infection,
16	KEGG_PATHWAY	UP_KEYWORDS	Acetylation, Alternative splicing, Antiviral defense, ATP-binding, Complete proteome, Cytoplasm, Direct protein sequencing, Endoplasmic reticulum, Glycoprotein, Immunity, Innate immunity, Lipoprotein, Magnesium, Metal-binding, Microsome, Mitochondrion, Myristate, Nucleotide-binding, Nucleotidyltransferase, Nucleus, Proteomics identification, Reference proteome, Repeat, RNA-binding, Transferase,
17	KEGG_PATHWAY	UP_SEQ_FEATURE	chain:2'-5'-oligoadenylate synthetase 2, compositionally biased region:Pro-rich (linker), lipid moiety-binding region:N-myristoyl glycine, modified residue, mutagenesis site, region of interest:OAS domain 1, region of interest:OAS domain 2, sequence conflict, splice variant,
18	KEGG_PATHWAY	ENSG00000111331	Download File
19	KEGG_PATHWAY	GOTERM_BP_DIRECT	nucleobase-containing compound metabolic process, immune response, response to virus, negative regulation of viral genome replication, defense response to virus, interferon-gamma-mediated signaling pathway, type I interferon signaling pathway, regulation of ribonuclease activity,
20	KEGG_PATHWAY	GOTERM_CC_DIRECT	extracellular space, nucleoplasm, cytoplasm, cytosol, plasma membrane, integral component of membrane, intracellular membrane-bound organelle,
21	KEGG_PATHWAY	GOTERM_MF_DIRECT	2'-5'-oligoadenylate synthetase activity, double-stranded RNA binding, protein binding, ATP binding, transferase activity, nucleotidyltransferase activity, metal ion binding,
22	KEGG_PATHWAY	INTERPRO	Nucleotidyl transferase domain, 2'-5'-oligoadenylate synthetase, N-terminal, 2'-5'-oligoadenylate synthetase, conserved site, 2'-5'-oligoadenylate synthetase 1, domain 2/C-terminal, 2'-5'-oligoadenylate synthetase,
23	KEGG_PATHWAY	KEGG_PATHWAY	Hepatitis C, Measles, Influenza A, Herpes simplex infection,
24	KEGG_PATHWAY	UP_KEYWORDS	3D-structure, Acetylation, Antiviral defense, ATP-binding, Complete proteome, Cytoplasm, Direct protein sequencing, Immunity, Innate immunity, Magnesium, Membrane, Metal-binding, Nucleotide-binding, Nucleotidyltransferase, Nucleus, Phosphoprotein, Polymorphism, Proteomics identification, Reference proteome, Repeat, RNA-binding, Transferase, Transmembrane, Transmembrane helix,
25	KEGG_PATHWAY	UP_SEQ_FEATURE	chain:2'-5'-oligoadenylate synthetase 3, modified residue, region of interest:Linker, region of interest:OAS domain 1, region of interest:OAS domain 2, region of interest:OAS domain 3, sequence conflict, sequence variant,
26	KEGG_PATHWAY	ENSG00000112667	Download File
27	KEGG_PATHWAY	GOTERM_BP_DIRECT	cell proliferation, nucleoside metabolic process, nucleotide metabolic process, deoxyribonucleoside monophosphate

Download File

	Kappa
	1.00
	0.84
	0.75
	0.71
	0.71
	0.68
	0.66
	0.59
	0.58
	0.57
	0.53
	0.51
	0.44
	0.41
	0.39
	0.38
	0.36
	0.32
	0.31



Warning: Your list contains more than 3000 genes. Please select a list with less than 3000 genes to use this tool.

Functional Annotation Clustering

[Help and Manual](#)

Functional Re

Options

[Rerun using options](#)

21943 term(s) wer

Similarity Score:

#	
1	KEGG_PATHWAY
2	KEGG_PATHWAY
3	KEGG_PATHWAY
4	KEGG_PATHWAY
5	KEGG_PATHWAY
6	UP_KEYWORDS
7	UP_KEYWORDS
8	GOTERM_BP_DI
9	GOTERM_CC_DI
10	GOTERM_MF_DI
11	UP_KEYWORDS
12	GOTERM_BP_DI
13	UP_KEYWORDS
14	UP_KEYWORDS
15	GOTERM_CC_DI
16	KEGG_PATHWAY
17	KEGG_PATHWAY
18	GOTERM_BP_DI
19	GOTERM_BP_DI

[!\[\]\(2cd9c7088d78b8430df8abb2365259bf_img.jpg\) Download File](#)

)

	Kappa
	1.00
	0.84
	0.75
	0.71
	0.71
	0.68
	0.66
	0.59
	0.58
	0.57
	0.53
	0.51
	0.44
	0.41
	0.39
	0.38
	0.36
	0.32
	0.31

G:Profiler

Nucleic Acids
Research

g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments

Jüri Reimand¹, Meelis Kull^{1,2,3}, Hedi Peterson^{2,3}, Jaanus Hansen¹ and Jaak Vilo^{1,2,3,*}

28 March, 2007 376 citations

- Ease of use
- Informative visual presentation of the results
- Supports ranked gene lists
- Works with most gene identifiers (even mixed)
- Convert gene IDs
- Find orthologous genes from other species
- Supports probe identifier for 31 species

G:Profiler

4 modules

G:PROFILER core

Identify pathways or biological processes enriched in the provided list

GO, KEGG, Reactome and TRANSFAC



Enriched pathways ranked by significance

Size of the query

Number of genes in the pathway

Ordered list

incrementally probes all possible sizes of the list and determine functional annotations and cut-points



Register how the pvalues change while increasing the size of the list of genes

G:Profiler

4 modules

G:PROFILER core

Identify pathways or biological processes enriched in the provided list
GO, KEGG, Reactome and TRANSFAC

↓
Enriched pathways ranked by significance
Size of the query
Number of genes in the pathway

Ordered list

incrementally probes all possible sizes of the list and determine functional annotations and cut-points

↓
Register how the pvalues change while increasing the size of the list of genes

G:SORTER

Gene expression similarity-based analysis

↓
Similarly or oppositely expressed genes from GEO datasets

↓
Obtain the “n” genes that are more similar to gene “x”

G:Profiler

4 modules

G:PROFILER core

Identify pathways or biological processes enriched in the provided list
GO, KEGG, Reactome and TRANSFAC

Enriched pathways ranked by significance
Size of the query
Number of genes in the pathway

Ordered list

incrementally probes all possible sizes of the list and determine functional annotations and cut-points

Register how the pvalues change while increasing the size of the list of genes

G:SORTER

Gene expression similarity-based analysis

Similarly or oppositely expressed genes from GEO datasets

Obtain the "n" genes that are more similar to gene "x"

Input: 1 gene ID

Select expression dataset
Correlation measure
Size of the list of genes

Enrichment of the list of genes

G:Profiler

4 modules

G:PROFILER core

Identify pathways or biological processes enriched in the provided list

*GO, KEGG, Reactome and
TRANSFAC*

Enriched pathways ranked by
significance
Size of the query
Number of genes in the
pathway

Ordered list

incrementally probes all possible sizes of the list and determine functional annotations and cut-points

Register how the pvalues change while increasing the size of the list of genes

G:SORTER

Gene expression similarity-based analysis

Similarly or oppositely expressed genes from GEO datasets

Obtain the “n” genes that are more similar to gene “x”

G:CONVERT

Convert gene IDs

List of gene names, Ids, proteins or microarray probes

Tabular file with the corresponding names and Ids for each gene

short description of the gene

G:ORTH

Identify orthologous genes
from other species

Data is downloaded from
Ensembl database

Input: List of genes
Select target organism
Tabular file with the
corresponding orthologs

Table 1. Overview of the functionality and data sources for different organisms in g-Profiler. Entries with (1) have less than 10 000 related GO associations

	<i>X. tropicola</i>	+	+	++	+	+
<i>T. rubripes</i>	+	+	++	++	+	+
<i>T. nigroviridis</i>	+	+	++			
<i>T. helvangeri</i>	+	+	++			
<i>P. troglodytes</i>	+	+	++		+	
<i>O. lampetra</i>	+	+	++			
<i>O. canadensis</i>	+	+	++			
<i>O. anatinus</i>	+	+	++			
<i>M. malapterus</i>	+	+	++			
<i>M. domesticus</i>	+	+	++			
<i>L. apodus</i>	+	+	++			
<i>G. gallus</i>	+	+	++	+		
<i>G. aculeatus</i>	+	+	++			
<i>F. carpio</i>	+	+	++			
<i>E. europaeus</i>	+	+	++			
<i>E. telegaster</i>	+	+	++			
<i>D. rerio</i>	+	+	++	+		
<i>D. noverca</i>	+	+	++			
<i>C. sarigyi</i>	+	+	++	+		
<i>C. porosus</i>	+	+	++			
<i>C. intestinalis</i>	+	+	++			
<i>C. familiaris</i>	+	+	++	+		
<i>B. taurus</i>	+	+	++	+		
<i>A. gambarus</i>	+	+	++			
<i>A. argeypti</i>	+	+	++	+		
<i>S. cerevisiae</i>	+	+	++	+		
<i>C. elegans</i>	+	+	++	+		
<i>D. melanogaster</i>	+	+	++	+		
<i>M. musculus</i>	+	+	++	+		
<i>H. sapiens</i>	+	+	++	+		

G:Profiler

g:Profiler

Juri.Reimand[@]ut.ee

BIIT 2005-2007

g:Data Gene Group Functional Profiling
g:Cocoa Compact Compare of Annotations
g:Convert Gene ID Converter
g:Sorter Expression Similarity Search
g:Orth Orthology search

Organism: Mus musculus

Gene query: CXCL1 98988_at_SELE

Significance threshold: SCS threshold, Bonferroni correction, Benjamini-Hochberg FDR

User p-value: 0.0001

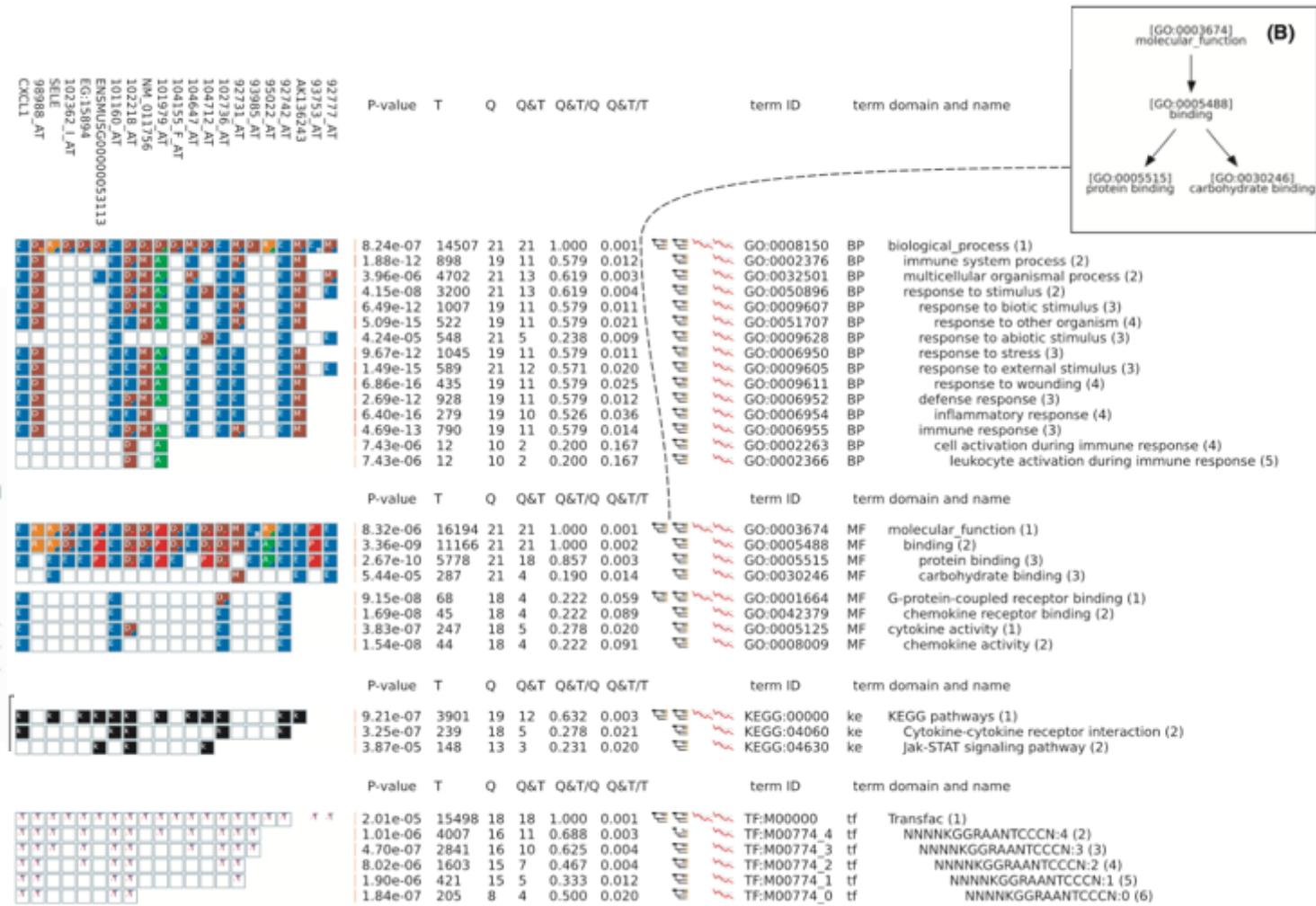
Output options: Significant only, Hierarchical sorting

Output type: Graphical (PNG)

Input options: Ordered query, Skip unknown entries in query

g:Profile! reset

g:Profiler home



G:Profiler

Nucleic Acids
Research

g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments

Jüri Reimand¹, Meelis Kull^{1,2,3}, Hedi Peterson^{2,3}, Jaanus Hansen¹ and Jaak Vilo^{1,2,3,*}

28 March, 2007 376 citations

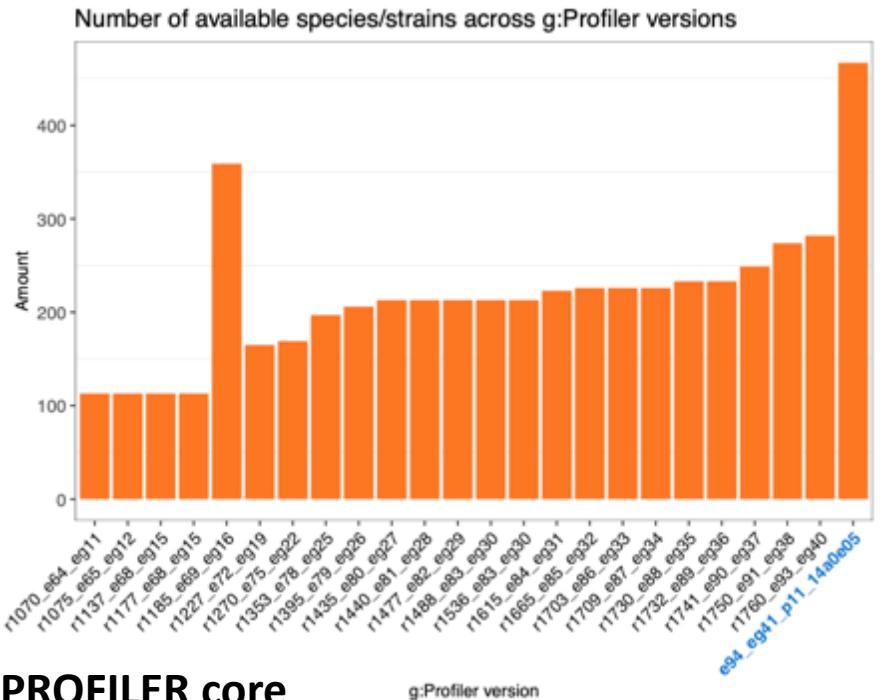
Nucleic Acids
Research

g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)

Uku Raudvere  ^{1,†}, Liis Kolberg  ^{1,†}, Ivan Kuzmin  ¹, Tambet Arak¹, Priit Adler^{1,2}, Hedi Peterson  ^{1,2,*} and Jaak Vilo  ^{1,2,3,*}

29 April, 2007 251 citations

G:Profiler



G:PROFILER core

↓

g:GOSt

Identify terms enriched in the provided list

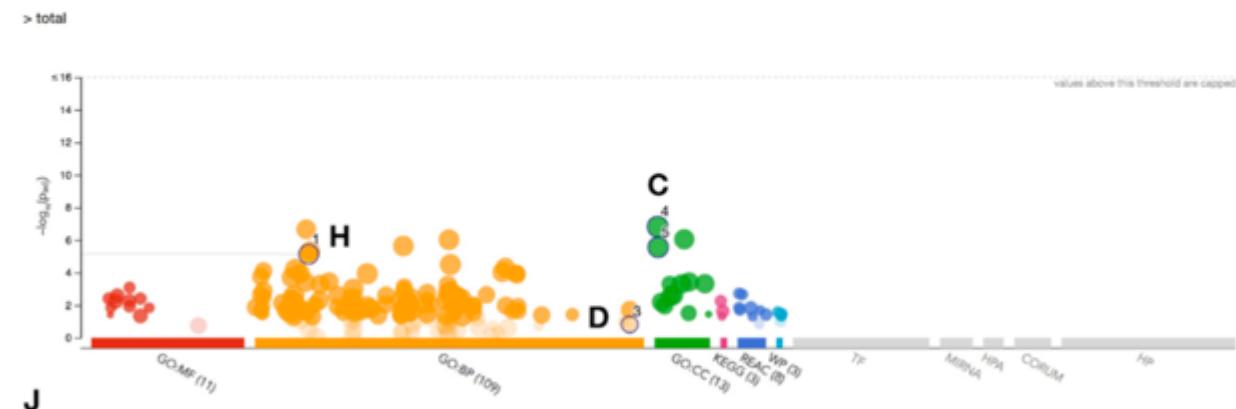
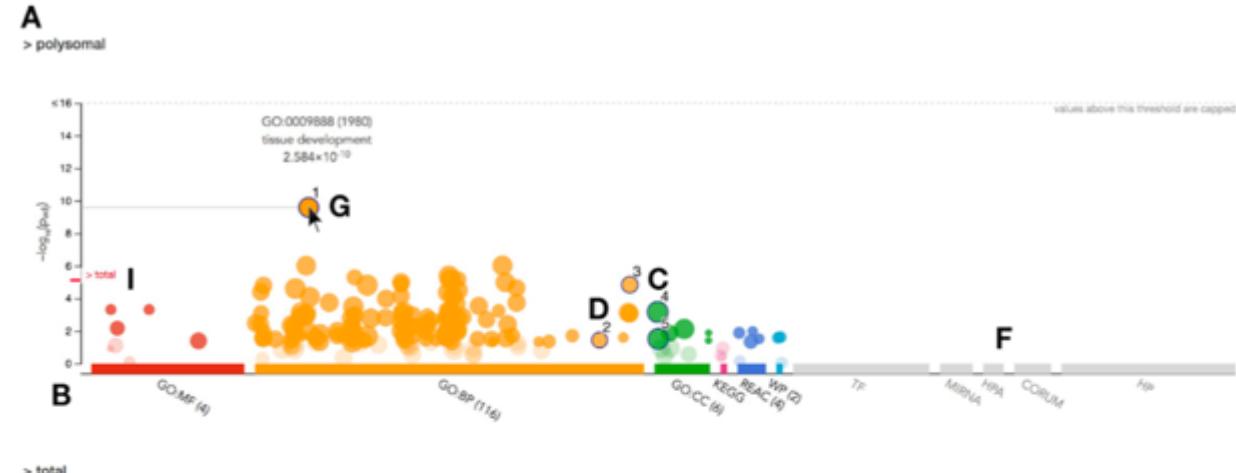
GO, KEGG, Reactome and TRANSFAC, WikiPathways, miRNA targets (miRTar-Base), tissue-specificity (Human Protein Atlas) and Human Phenotype Ontology

↓

g:SNPense

Map SNP rs-codes to gene names

Obtain chromosomal coordinates and predict variant effects.



id	source	term id	term name	E p_adj (polysomal)	E p_adj (total)
1	GO:BP	GO:0009888	tissue development	2.584×10^{-10}	7.503×10^{-6}
2	GO:BP	GO:1904018	positive regulation of vasculature development	3.609×10^{-9}	1.600×10^{-5}
3	GO:BP	GO:000146	negative regulation of cell motility	1.443×10^{-8}	1.629×10^{-4}
4	GO:CC	GO:0005576	extracellular region	8.793×10^{-4}	1.511×10^{-7}
5	GO:CC	GO:0005615	extracellular space	9.946×10^{-3}	2.786×10^{-6}

version
date
organism

e94_eg41_p11_14a0e05
30/01/2019, 19:18:38
hsapiens

g:Profiler

G:Profiler



News Archives Beta API R client FAQ Docs Contact Cite g:Profiler Services using g:P List of organisms



g:Profiler has been updated with new data from Ensembl and WormBase ParaSite.

Show more...

[Close](#)

g:GOSt
Functional profiling

g:Convert
Gene ID conversion

g:Orth
Orthology search

g:SNPense
SNP id to gene name

Query

Upload query

Upload bed file

Input is whitespace-separated list of genes

ENSG00000156261
ENSG00000160991
ENSG00000172732
ENSG00000115317
ENSG00000154518
ENSG00000181090
ENSG0000008086
ENSG00000033100
ENSG00000089177
ENSG00000068400
ENSG00000069869
ENSG00000145216
ENSG00000183426
ENSG00000134759
ENSG00000132842

[Run query ...](#)

[random example](#)

[mixed query example](#)

Options

* Organism:

Homo sapiens (Human)

Ordered query

Run as multiquery

[Advanced options ▾](#)

[Data sources ▾](#)

[Bring your data \(Custom GMT\) ▾](#)

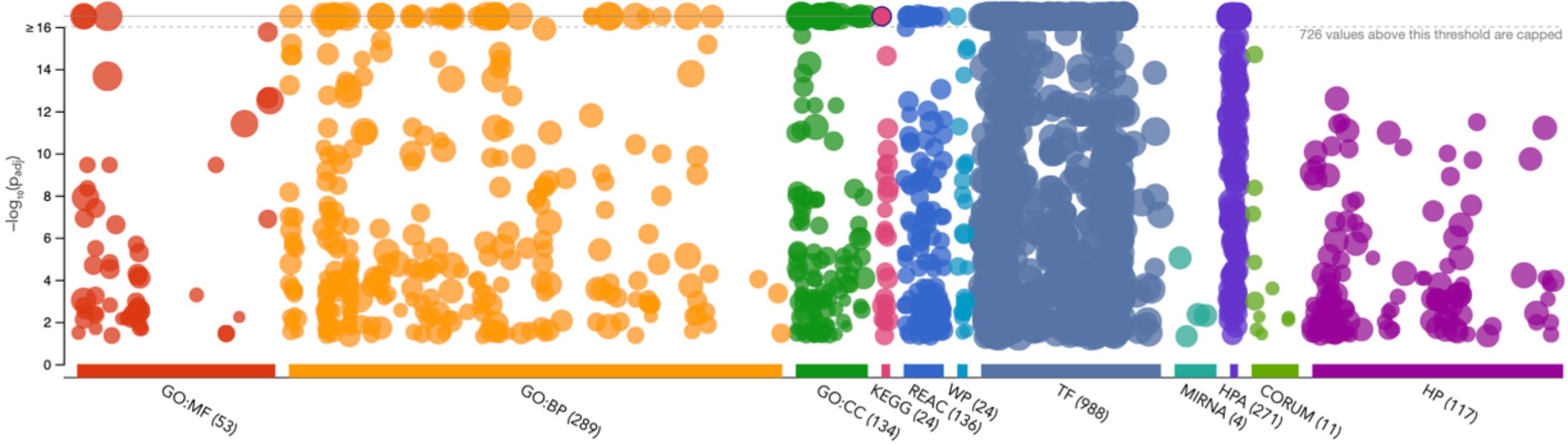


G:Profiler

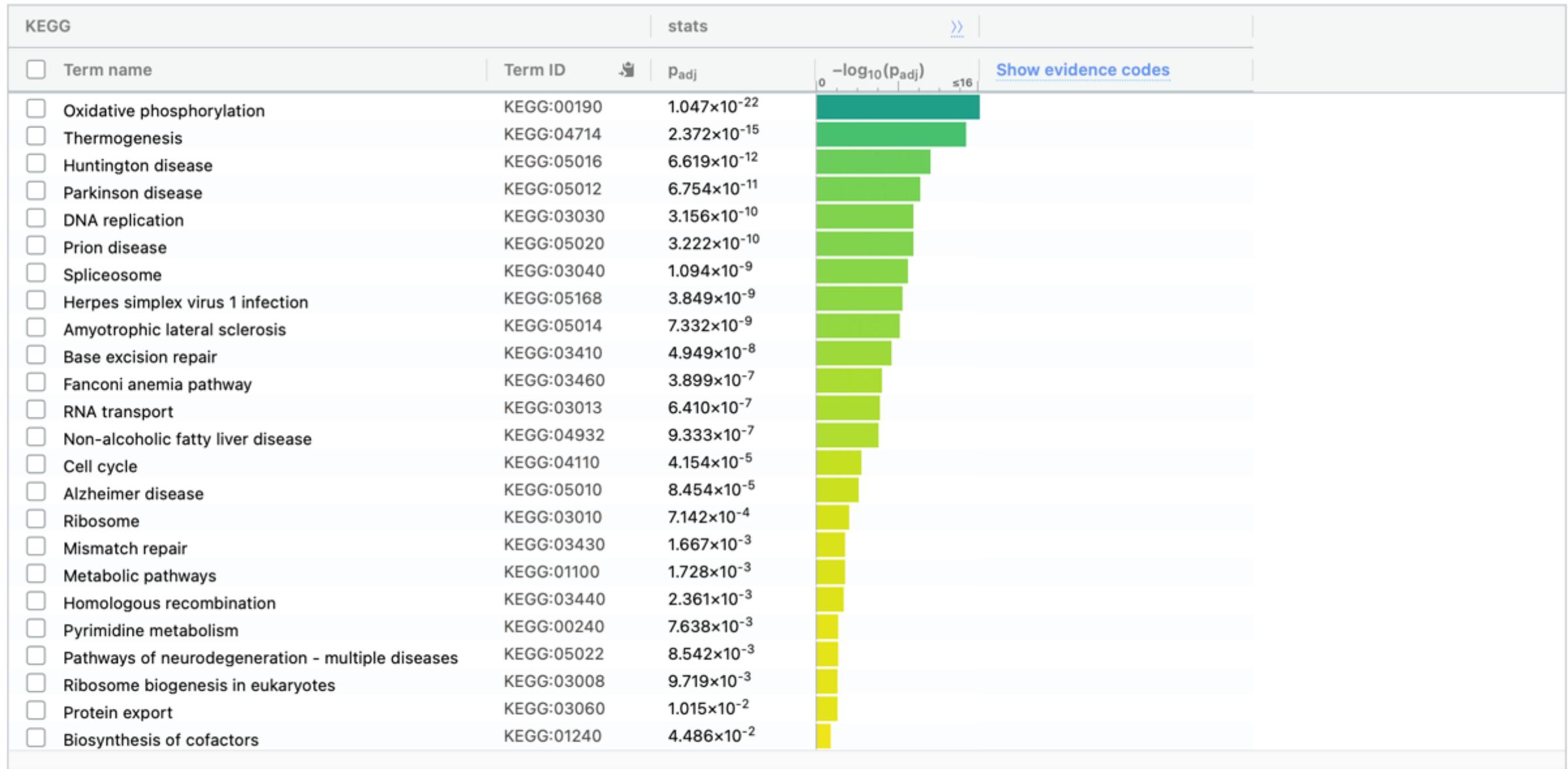
[Rerun query](#)[Ignore ambiguous query genes](#)[Permanently ignore ambiguous query genes](#)[Results](#)[Detailed Results](#)[Query info 153](#)[Export to PNG](#)[Show query URL](#)[Show short link](#) For print ? Capped ? Make unselected terms transparent

KEGG:00190 (133)
Oxidative phosphorylation
 1.047×10^{-22}

726 values above this threshold are capped



G:Profiler



G:Profiler

Package ‘gProfileR’

November 4, 2019

Version 0.7.0

License GPL (>= 2)

Description This package has been deprecated and will not be updated.

New users should use the package 'gprofiler2' (<<https://CRAN.R-project.org/package=gprofiler2>>)

for up-to-date data and improved functionality.

Functional enrichment analysis, gene identifier conversion and
mapping homologous genes across related organisms via the 'g:Profiler' toolkit
(<<https://biit.cs.ut.ee/gprofiler/>>).

```
gprofiler(query, organism = "hsapiens", sort_by_structure = T,  
ordered_query = F, significant = T, exclude_iea = F,  
underrep = F, evcodes = F, region_query = F, max_p_value = 1,  
min_set_size = 0, max_set_size = 0, min_isect_size = 0,  
correction_method = "analytical", hier_filtering = "none",  
domain_size = "annotated", custom_bg = "", numeric_ns = "",  
png_fn = NULL, include_graph = F, src_filter = NULL)
```

G:Profiler

```
> breastcup<-read.csv2("BreastCancer_genes_up.txt",stringsAsFactors = F,sep="\t",header=F)[,1]
> head(breastcup)
[1] "ENSG00000166803" "ENSG00000160957" "ENSG00000117724" "ENSG00000104889" "ENSG00000079462" "ENSG00000119333"
> length(breastcup)
[1] 4196
> background<-read.csv2("Breast_all_genes.txt",stringsAsFactors = F,sep="\t",header=F)[,1]
> head(background)
[1] "ENSG00000148773" "ENSG00000131747" "ENSG00000183856" "ENSG00000088325" "ENSG00000117724" "ENSG00000105664"
> length(background)
[1] 13110
> library("gProfileR")
> pathways<-gprofiler(breastcup,organism="hsapiens",custom_bg = background,src_filter = c("KEGG","REAC","GO"),
+ sort_by_structure = T,significant = T,correction_method = "gSCS") → 268 pathways
> pathways2<-gprofiler(breastcup,organism="hsapiens",custom_bg = background,src_filter = c("KEGG","REAC","GO"),→ 493 pathways
+ sort_by_structure = T,significant = T,correction_method = "fdr")
```

```
> reactome<-pathways[which(pathways[,10]=="rea"),c(9,12,3,6,4,5)]
> head(reactome[order(reactome[,3],decreasing = F),],10)
```

	term.id	term.name	p.value	overlap.size	term.size	query.size
187	REAC:R-HSA-1640170	Cell Cycle	2.97e-21	259	494	2224
201	REAC:R-HSA-69278	Cell Cycle, Mitotic	1.71e-17	216	411	2224
266	REAC:R-HSA-5389840	Mitochondrial translation elongation	9.42e-17	68	87	2224
267	REAC:R-HSA-5368286	Mitochondrial translation initiation	9.42e-17	68	87	2224
264	REAC:R-HSA-5368287	Mitochondrial translation	1.12e-15	70	93	2224
257	REAC:R-HSA-163200	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	5.13e-15	83	120	2224
265	REAC:R-HSA-5419276	Mitochondrial translation termination	5.23e-15	66	87	2224
258	REAC:R-HSA-611105	Respiratory electron transport	8.07e-12	68	99	2224
195	REAC:R-HSA-73886	Chromosome Maintenance	3.90e-11	48	62	2224
231	REAC:R-HSA-73894	DNA Repair	1.20e-10	138	262	2224

ToppGene Suite for gene list enrichment analysis and candidate gene prioritization

Jing Chen¹, Eric E. Bardes², Bruce J. Aronow^{2,3} and Anil G. Jegga^{2,3,*}

¹Department of Environmental Health, University of Cincinnati, ²Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center and ³Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

May 11, 2009 924 citations

Fuzzy Measures on the Gene Ontology for Gene Product Similarity

Mihail Popescu, James M. Keller, and Joyce A. Mitchell

Fuzzy measure

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A)g_\lambda(B)$$

for some $\lambda > -1$.

MTMR4

$G_1 = \{T_1 = 4721(\text{protein phosphatase activity}), T_2 = 6470(\text{protein amino acid dephosphorylation}), T_3 = 8270(\text{zinc ion binding})\}$

MTMR8

$G_2 = \{T_1 = 4721(\text{protein phosphatase activity}), T_2 = 6470(\text{protein amino acid dephosphorylation}), T_4 = 16787(\text{hydrolase activity})\}.$

$$P(T_k) = \frac{\text{Nº occurrences of term } T_k \text{ and its children}}{\text{Total Nº of terms}}$$

$$g_k = \frac{-\ln(p(T_k))}{\max\{-\ln(p(T_j))\}}$$

Densities for G_1 and G_2

$$\{g^{1i}\} = \{0.52, 0.57, 0.54\}$$

$$\{g^{2i}\} = \{0.52, 0.57, 0.33\}$$

$$(1 + \lambda) = \prod_{i=1}^n (1 + \lambda g^i).$$

$$G1 \lambda = -0.84,$$

$$G2 \lambda = -0.72,$$

$$g_1(\{T_1, T_2\}) = 0.84.$$

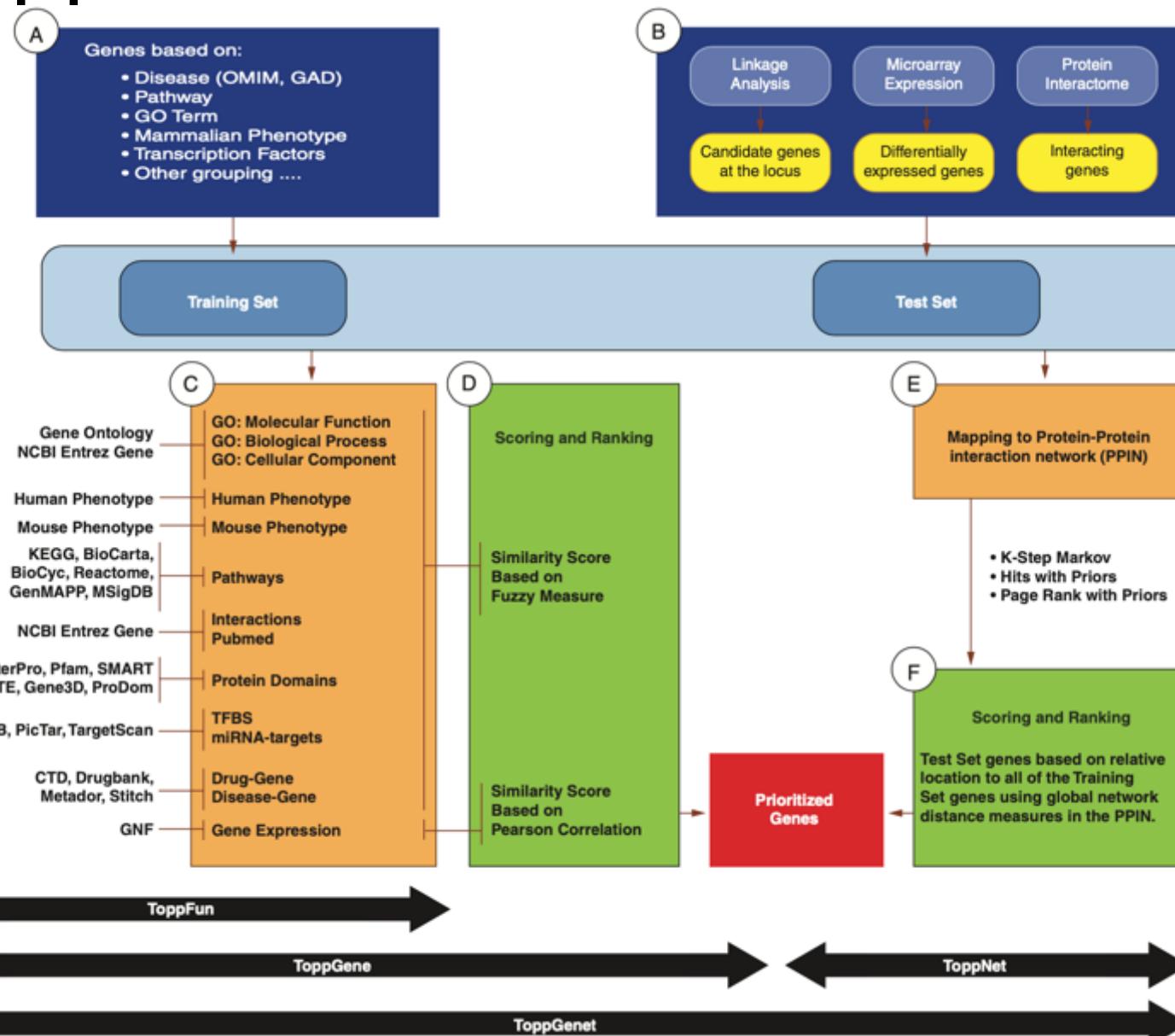
$$g_2(\{T_1, T_2\}) = 0.88$$

Fuzzy-based similarity measure: $s_{FMS}(G_1, G_2) = \frac{g_1(G_1 \cap G_2) + g_2(G_1 \cap G_2)}{2}$

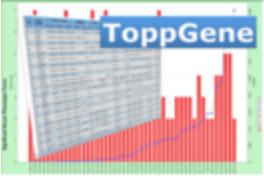
$$s_{FMS}(G_1, G_2) = \frac{g_1(\{T_1, T_2\}) + g_2(\{T_1, T_2\})}{2}$$

$$S_{FMS}(G_1, G_2) = 0.86$$

ToppGene



ToppGene



ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotations and protein interactions network

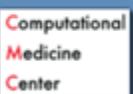
ToppGene:

- Home
- API
- Links
- Database details
- Supplementary
- Help
- Publications
- Terms of Use
- Contacts

Other Projects:

- ToppCluster
- Lungmap Heatmaps

Supported by:



- **ToppFun:** Transcriptome, ontology, phenotype, proteome, and pharmacome annotations based gene list functional enrichment analysis

Detect functional enrichment of your gene list based on Transcriptome, Proteome, Regulome (TFBS and miRNA), Ontologies (GO, Pathway), Phenotype (human disease and mouse phenotype), Pharmacome (Drug-Gene associations), literature co-citation, and other features.

- **ToppGene:** Candidate gene prioritization

Prioritize or rank candidate genes based on functional similarity to training gene list.

- **ToppNet:** Relative importance of candidate genes in networks

Prioritize or rank candidate genes based on topological features in protein-protein interaction network.

- **ToppGenet:** Prioritization of neighboring genes in protein-protein interaction network

Identify and prioritize the neighboring genes of the seeds in protein-protein interaction network based on functional similarity to the "seed" list (ToppGene) or topological features in protein-protein interaction network (ToppNet).

ToppGene



ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotations and protein interactions network

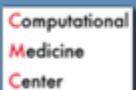
ToppGene:

- Home
- API
- Links
- Database details
- Supplementary
- Help
- Publications
- Terms of Use
- Contacts

Other Projects:

- ToppCluster
- Lungmap Heatmaps

Supported by:



ToppFun: Transcriptome, ontology, phenotype, proteome, and pharmacome annotations based gene list functional enrichment analysis

Select your gene identifier type, paste your sets below or select example set, then submit.

Symbol Type:

Ensembl ID

Background Symbol Type:

Ensembl ID

Example gene sets:

HGNC Symbol Entrez ID

(click on "HGNC Symbol" or "Entrez ID" to use the example training and test set of genes)

Enrichment Gene Set:

ENSG00000165632
ENSG0000006282
ENSG00000145592
ENSG00000115415
ENSG00000170606
ENSG00000156261
ENSG00000160991
ENSG00000172732
ENSG00000115317
ENSG00000154518
ENSG00000181090
ENSG0000008086
ENSG00000033100
ENSG00000089177
ENSG00000068400
ENSG00000069869
ENSG00000145216
ENSG00000183426
ENSG00000134759
ENSG00000132842

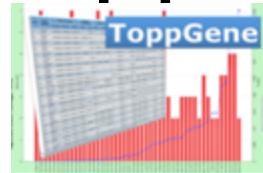
Background Gene Set:

ENSG00000125740
ENSG00000175445
ENSG0000004799
ENSG00000186642
ENSG00000196616
ENSG00000131386
ENSG00000134339
ENSG00000079435
ENSG00000135218
ENSG00000181092
ENSG000000101938
ENSG00000165269
ENSG00000168477
ENSG00000187134
ENSG00000167676
ENSG00000151632
ENSG00000170323
ENSG00000173432
ENSG00000166819
ENSG00000162896

Clear

Submit

ToppGene



ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotations and protein interactions network

ToppGene:

- Home
- API
- Links
- Database details
- Supplementary
- Help
- Publications
- Terms of Use
- Contacts

Other Projects:

- ToppCluster
- Lungmap Heatmaps

Supported by:



Input Gene List (4032 / 4196)

Entered	Human Symbol	Gene ID
ENSG00000166803 PCLAF	(PCNA clamp associated factor)	9768
ENSG00000160957 RECQL4	(RecQ like helicase 4)	9401
ENSG00000117724 CENPF	(centromere protein F)	1063
ENSG00000104889 RNASEH2A	(ribonuclease H2 subunit A)	10535
ENSG00000079462 PAFAH1B3	(platelet activating factor acetylhydrolase 1b catalytic subunit 3)	5050
ENSG00000119333 DYNC2I2	(dynein 2 intermediate chain 2)	89891
ENSG0000013810 TACC3	(transforming acidic coiled-coil containing protein 3)	10460
ENSG00000101412 E2F1	(E2F transcription factor 1)	1869
ENSG00000111206 FOXM1	(forkhead box M1)	2305
ENSG00000160298 C21orf58	(chromosome 21 open reading frame 58)	54058
ENSG00000123219 CENPK	(centromere protein K)	64105
ENSG00000076382 SPAG5	(sperm associated antigen 5)	10615
ENSG00000166851 PLK1	(polo like kinase 1)	5347
ENSG00000167900 TK1	(thymidine kinase 1)	7083
ENSG00000113810 SMC4	(structural maintenance of chromosomes 4)	10051
ENSG00000134222 PSRC1	(proline and serine rich coiled-coil 1)	84722
ENSG00000140525 FANCI	(FA complementation group I)	55215
ENSG00000137804 NUSAP1	(nucleolar and spindle associated protein 1)	51203
ENSG00000148773 MKI67	(marker of proliferation Ki-67)	4288
ENSG00000137807 KIF23	(kinesin family member 23)	9493
ENSG00000122952 ZWINT	(ZW10 interacting kinetochore protein)	11130

Genes Not Found

Entered	Status
ENSG00000153406	Duplicated
ENSG00000278243	Duplicated
ENSG00000275911	Duplicated
ENSG00000276657	Duplicated
ENSG00000282228	Not Found
ENSG00000240972	Duplicated
ENSG00000280755	Not Found
ENSG00000280529	Not Found
ENSG00000283018	Not Found
ENSG00000188092	Duplicated
ENSG00000282685	Not Found
ENSG00000274829	Duplicated
ENSG00000282941	Not Found
ENSG00000182973	Duplicated
ENSG00000281024	Not Found
ENSG00000274287	Duplicated
ENSG00000186409	Duplicated
ENSG00000274891	Not Found

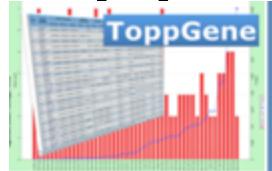
Background Gene List (12417 / 13110)

Entered	Human Symbol	Gene ID
ENSG00000148773 MKI67	(marker of proliferation Ki-67)	4288
ENSG00000131747 TOP2A	(DNA topoisomerase II alpha)	7153
ENSG00000183856 IQGAP3	(IQ motif containing GTPase activating protein 3)	128239
ENSG00000088325 TPX2	(TPX2 microtubule nucleation factor)	22974
ENSG00000117724 CENPF	(centromere protein F)	1063
ENSG00000105664 COMP	(cartilage oligomeric matrix protein)	1311
ENSG00000138778 CENPE	(centromere protein E)	1062
ENSG00000167900 TK1	(thymidine kinase 1)	7083
ENSG00000122952 ZWINT	(ZW10 interacting kinetochore protein)	11130
ENSG00000166851 PLK1	(polo like kinase 1)	5347
ENSG00000092853 CLSPN	(claspin)	63967
ENSG00000258947 TUBB3	(tubulin beta 3 class III)	10381
ENSG00000101412 E2F1	(E2F transcription factor 1)	1869
ENSG00000126709 IFI6	(interferon alpha inducible protein 6)	2537
ENSG00000115414 FN1	(fibronectin 1)	2335
ENSG00000189057 FAM111B	(family with sequence similarity 111 member B)	374393
ENSG00000111206 FOXM1	(forkhead box M1)	2305
ENSG00000276043 UHRF1	(ubiquitin like with PHD and ring finger domains 1)	29128
ENSG00000117122 MFAP2	(microfibril associated protein 2)	4237
ENSG00000134057 CCNB1	(cyclin B1)	891
ENSG00000075218 GTSE1	(G2 and S-phase expressed 1)	51512

Genes Not Found

Entered	Status
ENSG00000280755	Not Found
ENSG00000240972	Duplicated
ENSG00000282941	Not Found
ENSG00000282685	Not Found
ENSG00000278243	Duplicated
ENSG00000238134	Duplicated
ENSG00000281230	Not Found
ENSG00000204839	Duplicated
ENSG00000282928	Not Found
ENSG00000153406	Duplicated
ENSG00000276657	Duplicated
ENSG00000180921	Duplicated
ENSG00000282228	Not Found
ENSG00000165949	Duplicated
ENSG00000280529	Not Found
ENSG00000224736	Not Found
ENSG00000186409	Duplicated
ENSG00000234078	Duplicated

ToppGene



ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotation and clinical variants

pValue Method: Probability density function

ToppGene

Results

[Go To Start Page](#)

Input Parameters [Show Detail]

Training Results [Expand All] [Download All] [Sparse Matrix] Display pValues and Scores as Scientific (4 significant digits) Table row limit 50 [?]

1: GO: Molecular Function [Display Chart] 3614 input genes in category / 2501 annotations before applied cutoff / 11348 genes in category

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
1	GO:0003723 RNA binding		5.435E-10	1.359E-6	1.142E-5	1.359E-6	533	1359
2	GO:0003954 NADH dehydrogenase activity		1.578E-7	9.869E-5	8.292E-4	3.948E-4	30	42
3	GO:0050136 NADH dehydrogenase (quinone) activity		1.578E-7	9.869E-5	8.292E-4	3.948E-4	30	42
4	GO:0008137 NADH dehydrogenase (ubiquinone) activity		1.578E-7	9.869E-5	8.292E-4	3.948E-4	30	42
5	GO:0140097 catalytic activity, acting on DNA		2.419E-7	1.210E-4	1.017E-3	6.051E-4	133	289

[Show 24 more annotations](#)

2: GO: Biological Process [Display Chart] 3575 input genes in category / 10213 annotations before applied cutoff / 11307 genes in category

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
1	GO:0070125 mitochondrial translational elongation		3.124E-18	3.190E-14	3.129E-13	3.190E-14	67	87
2	GO:0070126 mitochondrial translational termination		2.213E-17	1.130E-13	1.108E-12	2.260E-13	66	87
3	GO:0032543 mitochondrial translation		4.895E-17	1.666E-13	1.634E-12	4.999E-13	88	131
4	GO:0140053 mitochondrial gene expression		1.368E-15	3.494E-12	3.427E-11	1.398E-11	98	157
5	GO:0006259 DNA metabolic process		2.320E-15	4.739E-12	4.649E-11	2.370E-11	355	799

[Show 45 more annotations](#)

3: GO: Cellular Component [Display Chart] 3616 input genes in category / 1377 annotations before applied cutoff / 11436 genes in category

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
1	GO:0098798 mitochondrial protein complex		2.196E-26	3.024E-23	2.360E-22	3.024E-23	160	250
2	GO:0005743 mitochondrial inner membrane		6.499E-24	4.475E-21	3.493E-20	8.950E-21	246	456
3	GO:0019866 organelle inner membrane		1.963E-22	9.012E-20	7.034E-19	2.704E-19	259	496
4	GO:0005761 mitochondrial ribosome		3.728E-18	1.027E-15	8.014E-15	5.134E-15	68	89
5	GO:0000313 organellar ribosome		3.728E-18	1.027E-15	8.014E-15	5.134E-15	68	89

[Show 45 more annotations](#)

4: Human Phenotype [Display Chart] 841 input genes in category / 5830 annotations before applied cutoff / 2920 genes in category

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
1	HP:0025454 Abnormal CSF metabolite level		4.490E-9	2.618E-5	2.421E-4	2.618E-5	47	78
2	HP:0002490 Increased CSF lactate		2.534E-8	4.924E-5	4.554E-4	1.477E-4	44	74
3	HP:0030085 Abnormal CSF lactate level		2.534E-8	4.924E-5	4.554E-4	1.477E-4	44	74
4	HP:0000340 Sloping forehead		3.441E-8	5.015E-5	4.638E-4	2.006E-4	48	84
5	HP:0007874 Almond-shaped palpebral fissure		1.663E-7	1.616E-4	1.495E-3	9.698E-4	16	18

[Show 35 more annotations](#)

ToppGene

6: Domain [Display Chart] 3580 input genes in category / 9321 annotations before applied cutoff / 11338 genes in category

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation	
1	IPR010920	LSM dorm	InterPro	1.190E-6	4.780E-3	4.645E-2	1.109E-2	17	20
2	SM00651	Sm	SMART	2.051E-6	4.780E-3	4.645E-2	1.912E-2	15	17
3	PF01423	LSM	Pfam	2.051E-6	4.780E-3	4.645E-2	1.912E-2	15	17
4	IPR001163	LSM dom euk/arc	InterPro	2.051E-6	4.780E-3	4.645E-2	1.912E-2	15	17
5	PS51808	CHCH	PROSITE	1.363E-5	2.541E-2	2.469E-1	1.270E-1	17	22

7: Pathway [Display Chart] 2382 input genes in category / 3286 annotations before applied cutoff / 7578 genes in category

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation	
1	1269741	Cell Cycle	BioSystems: REACTOME	9.301E-23	3.056E-19	2.651E-18	3.056E-19	244	463
2	1268844	Mitochondrial translation elongation	BioSystems: REACTOME	1.978E-19	1.894E-16	1.643E-15	6.501E-16	67	85
3	1268843	Mitochondrial translation initiation	BioSystems: REACTOME	1.978E-19	1.894E-16	1.643E-15	6.501E-16	67	85
4	1269763	Cell Cycle, Mitotic	BioSystems: REACTOME	2.306E-19	1.894E-16	1.643E-15	7.576E-16	205	388
5	1268842	Mitochondrial translation	BioSystems: REACTOME	2.574E-18	1.692E-15	1.467E-14	8.458E-15	69	91
6	1268845	Mitochondrial translation termination	BioSystems: REACTOME	1.173E-17	6.422E-15	5.571E-14	3.853E-14	65	85
7	1268838	Organelle biogenesis and maintenance	BioSystems: REACTOME	2.157E-17	1.013E-14	8.786E-14	7.089E-14	165	303
8	1270127	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	BioSystems: REACTOME	3.464E-16	1.423E-13	1.234E-12	1.138E-12	78	114
9	82942	Oxidative phosphorylation	BioSystems: KEGG	3.624E-15	1.323E-12	1.148E-11	1.191E-11	76	113
10	1269649	Gene Expression	BioSystems: REACTOME	5.834E-15	1.917E-12	1.663E-11	1.917E-11	593	1485
11	M19540	Oxidative phosphorylation	MSigDB C2 BIOCARTA (v7.1)	3.727E-14	1.113E-11	9.657E-11	1.225E-10	73	110
12	1269853	Chromosome Maintenance	BioSystems: REACTOME	8.174E-14	2.238E-11	1.942E-10	2.686E-10	47	60
13	1270128	Respiratory electron transport	BioSystems: REACTOME	5.162E-13	1.305E-10	1.132E-9	1.696E-9	63	93
14	1270350	DNA Repair	BioSystems: REACTOME	1.536E-12	3.606E-10	3.128E-9	5.048E-9	126	237
15	1270121	The citric acid (TCA) cycle and respiratory electron transport	BioSystems: REACTOME	9.539E-12	2.090E-9	1.813E-8	3.135E-8	88	152
16	1268847	Anchoring of the basal body to the plasma membrane	BioSystems: REACTOME	4.980E-11	1.023E-8	8.872E-8	1.636E-7	58	89
17	1269875	DNA Replication	BioSystems: REACTOME	1.874E-10	3.522E-8	3.055E-7	6.156E-7	58	91
18	M7272	Parkinson's disease	MSigDB C2 BIOCARTA (v7.1)	1.929E-10	3.522E-8	3.055E-7	6.340E-7	64	104
19	1309099	Homology Directed Repair	BioSystems: REACTOME	2.204E-10	3.812E-8	3.307E-7	7.244E-7	56	87
20	1309095	DNA Double-Strand Break Repair	BioSystems: REACTOME	2.430E-10	3.993E-8	3.464E-7	7.985E-7	67	111
21	1269854	Nucleosome assembly	BioSystems: REACTOME	2.804E-10	4.189E-8	3.634E-7	9.215E-7	25	28
22	1269855	Deposition of new CENPA-containing nucleosomes at the centromere	BioSystems: REACTOME	2.804E-10	4.189E-8	3.634E-7	9.215E-7	25	28
23	83098	Parkinson's disease	BioSystems: KEGG	1.067E-9	1.461E-7	1.267E-6	3.506E-6	68	116
24	1269753	G2/M Checkpoints	BioSystems: REACTOME	1.067E-9	1.461E-7	1.267E-6	3.506E-6	68	116
25	1339146	Complex I biogenesis	BioSystems: REACTOME	1.825E-9	2.398E-7	2.081E-6	5.996E-6	38	53
26	1269856	Telomere Maintenance	BioSystems: REACTOME	2.904E-9	3.670E-7	3.183E-6	9.541E-6	32	42
27	83097	Alzheimer's disease	BioSystems: KEGG	3.708E-9	4.513E-7	3.915E-6	1.219E-5	76	137
28	1268846	Cilium Assembly	BioSystems: REACTOME	4.258E-9	4.997E-7	4.335E-6	1.399E-5	86	161
29	1269784	DNA strand elongation	BioSystems: REACTOME	5.986E-9	6.783E-7	5.884E-6	1.967E-5	25	30
30	1269779	Synthesis of DNA	BioSystems: REACTOME	6.578E-9	7.205E-7	6.250E-6	2.161E-5	53	86
31	1269688	Processing of Capped Intron-Containing Pre-mRNA	BioSystems: REACTOME	7.634E-9	8.092E-7	7.019E-6	2.508E-5	109	219
32	1309100	HDR through Homologous Recombination (HR) or Single Strand Annealing (SSA)	BioSystems: REACTOME	8.107E-9	8.325E-7	7.222E-6	2.664E-5	51	82
33	1269742	Cell Cycle Checkpoints	BioSystems: REACTOME	8.828E-9	8.790E-7	7.625E-6	2.901E-5	76	139
34	1269777	S Phase	BioSystems: REACTOME	9.498E-9	9.180E-7	7.963E-6	3.121E-5	63	109
35	83039	DNA replication	BioSystems: KEGG	1.072E-8	9.788E-7	8.491E-6	3.524E-5	27	34
36	M16853	DNA replication	MSigDB C2 BIOCARTA (v7.1)	1.072E-8	9.788E-7	8.491E-6	3.524E-5	27	34
37	83100	Huntington's disease	BioSystems: KEGG	1.278E-8	1.135E-6	9.848E-6	4.201E-5	84	159
38	M16024	Alzheimer's disease	MSigDB C2 BIOCARTA (v7.1)	1.622E-8	1.403E-6	1.217E-5	5.331E-5	72	131

ToppGene

Results

[Go To Start Page](#)

Training Parameters [Show All]

Training Gene Subnetwork [Hide All] Display pValues and Scores as Scientific (4 significant digits) CSV

Number of genes: 15851
Number of interactions: 460507
Too many interactions to display

Format: XGMML (An XML based format compatible with Cytoscape) CSV

Test Genes [Hide All]

Rank	ID	Name	Interactant count	Score
1	25962	VIRMA	2897	3.120E-3
2	3191	HNRNPL	1735	2.995E-3
3	7706	TRIM25	2413	2.909E-3
4	351	APP	2286	2.905E-3
5	57664	PLEKHA4	2967	2.900E-3
6	2100	ESR2	2297	2.452E-3
7	1994	ELAVL1	1836	2.129E-3
8	3845	KRAS	1588	1.846E-3
9	4609	MYC	2000	1.765E-3
10	4914	NTRK1	1984	1.725E-3
11	328	APEX1	1024	1.563E-3
12	8359	H4C1	1561	1.454E-3
13	6047	RNF4	1339	1.403E-3
14	10482	NXF1	1193	1.327E-3
15	7182	NR2C2	1360	1.298E-3
16	112399	EGLN3	1285	1.239E-3
17	1956	EGFR	1320	1.235E-3
18	8726	EED	1364	1.207E-3
19	9343	EFTUD2	1332	1.163E-3
20	7157	TP53	1257	1.145E-3
21	10155	TRIM28	1059	1.090E-3
22	330	BIRC3	1328	1.087E-3
23	79155	TNIP2	940	1.038E-3
24	63891	RNF123	782	9.216E-4
25	55716	LMBR1L	942	9.093E-4
26	7415	VCP	982	8.991E-4
27	4000	LMNA	861	8.604E-4
28	4223	MEOX2	692	8.504E-4
29	3312	HSPA8	933	8.409E-4

ToppGene

Results

[Go To Start Page](#)

Training Parameters [Show All]

Training Gene Subnetwork [Hide All] Display pValues and Scores as Scientific (4 significant digits) XML

Number of genes: 15851
Number of interactions: 460507
Too many interactions to display

Format: XGMML (An XML based format compatible with Cytoscape) XML Export

Test Genes [Hide All]

Rank	ID	Name	Interactant count	Score
1	25962	VIRMA	2897	3.120E-3
2	3191	HNRNPL	1735	2.995E-3
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17	Jia-Yi Qian <small>1</small>	, Jian Gao <small>2</small>	, Xi Sun <small>1</small>	, Meng-Da Cao <small>3</small>
18				, Liang Shi <small>1</small>
19				, Tian-Song Xia <small>1</small>
20				, Wen-Bin Zhou <small>1</small>
21				, Shui Wang <small>4</small>
22				, Qiang Ding <small>5</small>
23				, Ji-Fu Wei <small>6</small>
24				
25				
26				
27				
28				
29				

KIAA1429 acts as an oncogenic factor in breast cancer by regulating CDK1 in an N6-methyladenosine-independent manner

Jia-Yi Qian 1, Jian Gao 2, Xi Sun 1, Meng-Da Cao 3, Liang Shi 1, Tian-Song Xia 1,
Wen-Bin Zhou 1, Shui Wang 4, Qiang Ding 5, Ji-Fu Wei 6

330	BIRC3	1328	1.087E-3
79155	TNIP2	940	1.038E-3
63891	RNF123	782	9.216E-4
55716	LMBR1L	942	9.093E-4
7415	VCP	982	8.991E-4
4000	LMNA	861	8.604E-4
4223	MEOX2	692	8.504E-4
3312	HSPA8	933	8.409E-4

Published: 15 June 2003

nature genetics

PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes

Vamsi K Mootha, Cecilia M Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, Emma Carlsson, Martin Ridderstråle, Esa Laurila, Nicholas Houstis, Mark J Daly, Nick Patterson, Jill P Mesirov, Todd R Golub, Pablo Tamayo, Bruce Spiegelman, Eric S Lander, Joel N Hirschhorn, David Altshuler & Leif C Groop

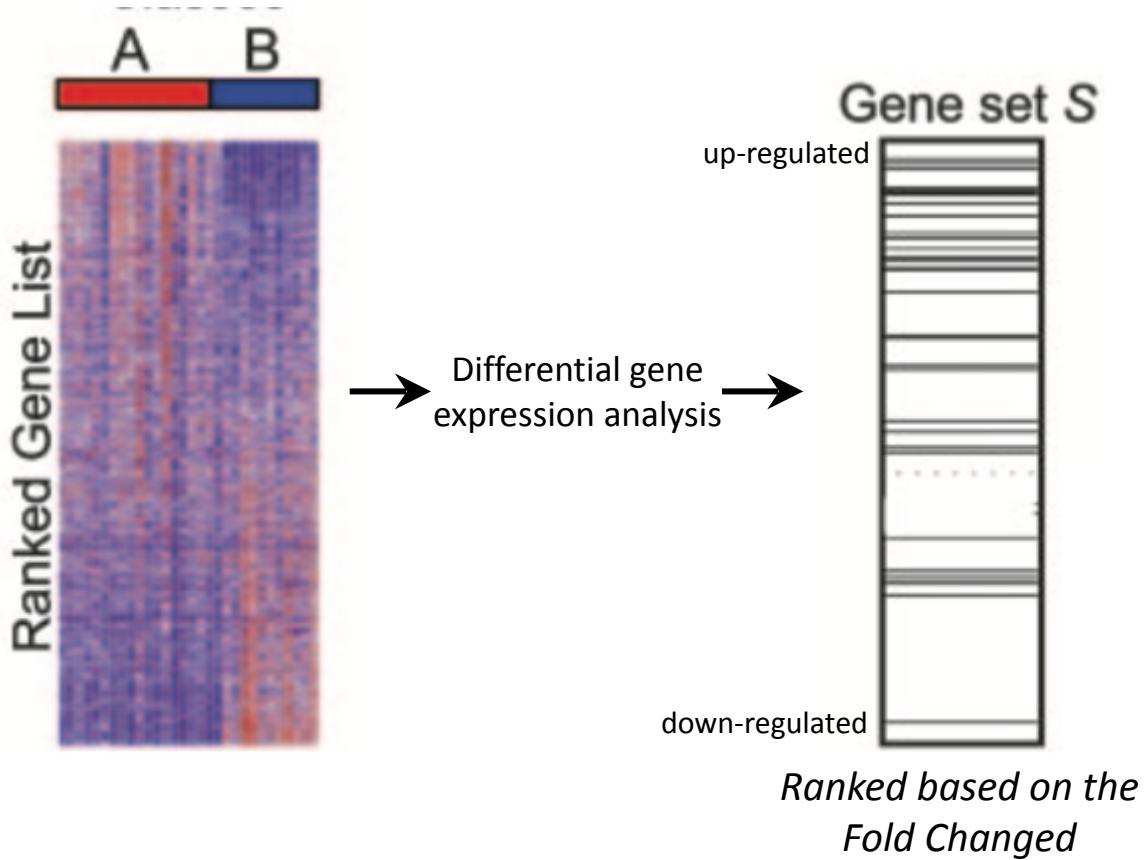


Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles

Aravind Subramanian^{a,b}, Pablo Tamayo^{a,b}, Vamsi K. Mootha^{a,c}, Sayan Mukherjee^d, Benjamin L. Ebert^{a,e}, Michael A. Gillette^{a,f}, Amanda Paulovich^g, Scott L. Pomeroy^h, Todd R. Golub^{a,e}, Eric S. Lander^{a,c,i,j,k}, and Jill P. Mesirova^{a,k}

August 2, 2005 13,302 citations

GSEA (*method*)



GSEA (method)

Gene set of interest ← → Correlation with the phenotype

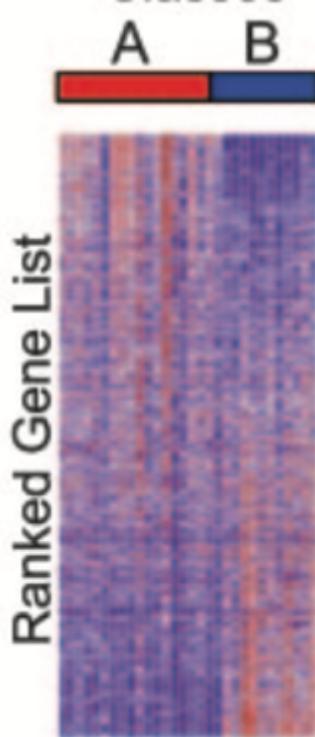
Position in the ranked list

$$P_{\text{hit}}(S, i) = \sum_{\substack{g_j \in S \\ j \leq i}} \frac{|r_j|^p}{N_R}, \quad \sum_{g_i \in S} |r_j|^p$$

$$\text{ES} = \max(P_{\text{hit}} - P_{\text{miss}})$$

$$P_{\text{miss}}(S, i) = \sum_{\substack{g_j \notin S \\ j \leq i}} \frac{1}{(N - N_H)} \cdot \frac{1}{\text{Number of genes in the gene set}}$$

Total number
of genes



Gene set S

up-regulated



down-regulated

Ranked based on the
Fold Changed

GSEA (method)

Gene set of interest ← → *Correlation with the phenotype*

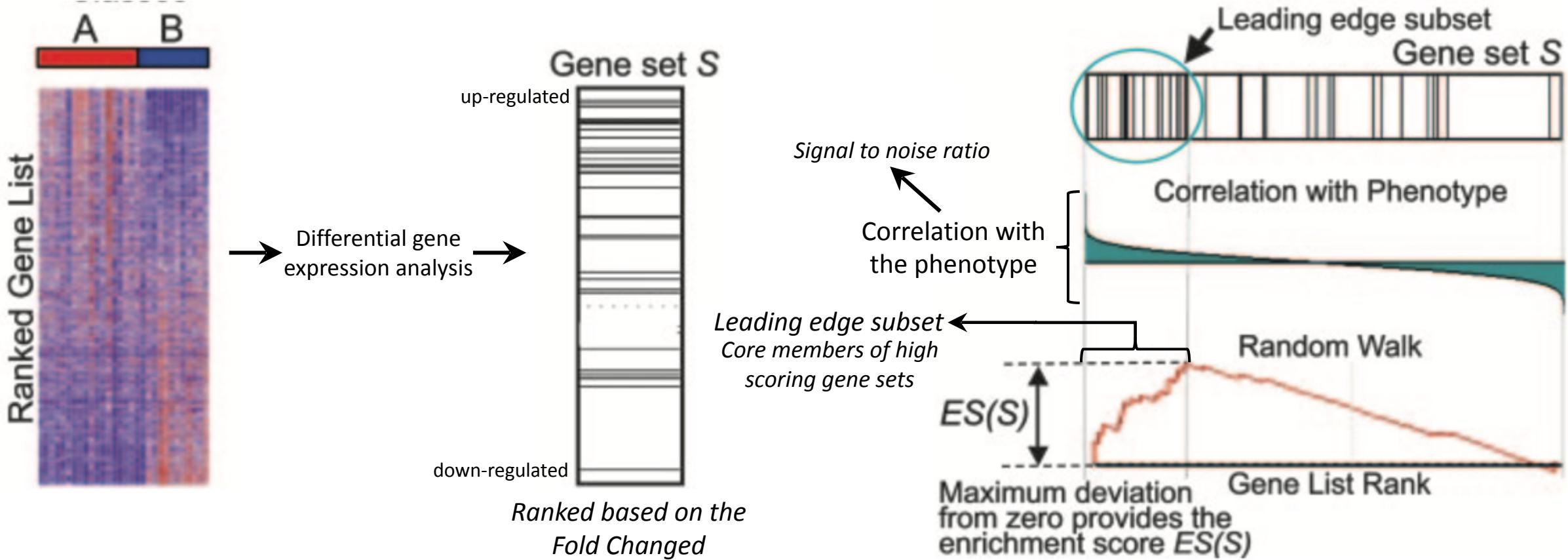
$$P_{\text{hit}}(S, i) = \sum_{\substack{g_j \in S \\ j \leq i}} \frac{|r_j|^p}{N_R}, \quad \sum_{g_i \in S} |r_j|^p$$

Position in the ranked list ←

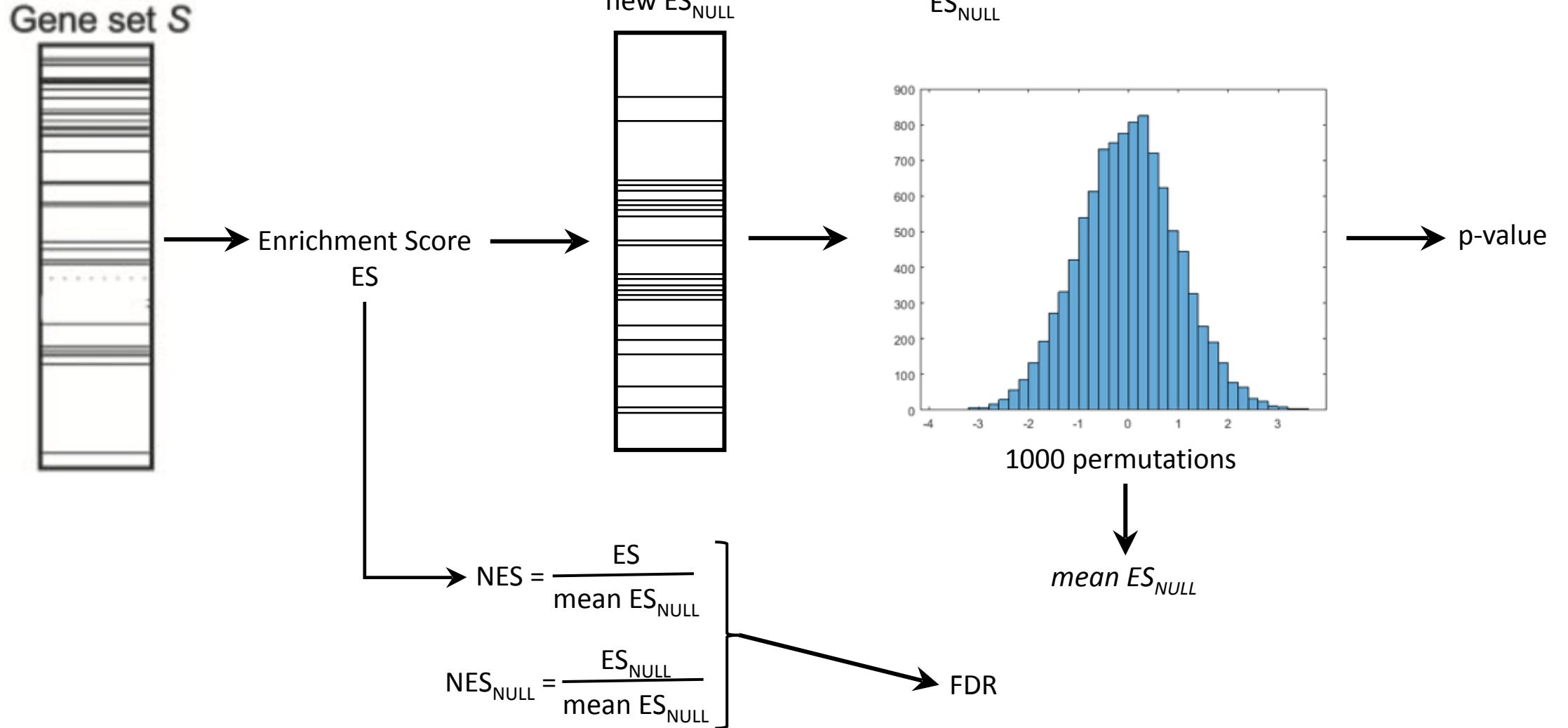
$$P_{\text{miss}}(S, i) = \sum_{\substack{g_j \notin S \\ j \leq i}} \frac{1}{(N - N_H)} \cdot \frac{1}{N_H} \quad \rightarrow \text{Number of genes in the gene set}$$

Total number of genes ←

$$\text{ES} = \max(P_{\text{hit}} - P_{\text{miss}})$$



GSEA (*method*)



GSEA

GSEA
Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact Team

<http://www.gsea-msigdb.org/gsea/index.jsp>

UC San Diego

BROAD
INSTITUTE

Overview

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether *a priori* defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes).

- ▶ [Download](#) the GSEA software and additional resources to analyze, annotate and interpret enrichment results.
- ▶ [Explore the Molecular Signatures Database \(MSigDB\)](#), a collection of annotated gene sets for use with GSEA software.
- ▶ [View documentation](#) describing GSEA and MSigDB.
- ▶ View guidelines for [using RNA-seq datasets with GSEA](#).
- ▶ Use the [GenePattern](#) platform to run analyses, including classical GSEA and a variation designed for single-sample analysis ([ssGSEA](#)).

What's New

19-Sep-2020: MSigDB 7.2 released. This release includes a substantial reorganization of C5 to accommodate the addition of the Human Phenotype Ontology, the addition of gene sets from WikiPathways to C2:CP, and the promotion of SCSig to C8, among other minor updates and additions. See the [release notes](#) for details.

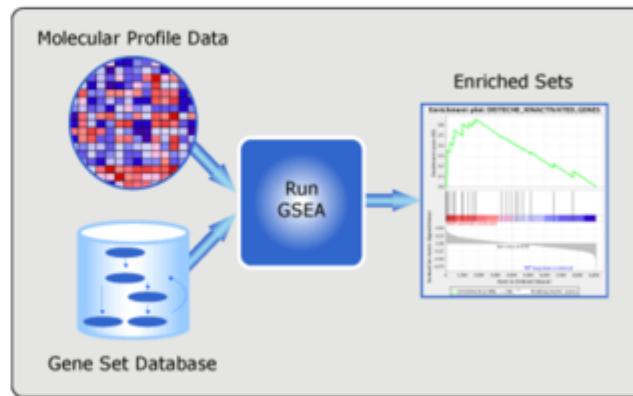
30-Jul-2020: GSEA 4.1.0 released. This release includes improved support for macOS Catalina, updated and improved Enrichment Reports, and numerous bug fixes. See the [release notes](#) for more information.

30-Mar-2020: MSigDB 7.1 released. New C3 sub-collections from the [Gene Transcription Regulation Database](#) and [miRDB](#). Updated gene sets from Reactome 72 and GO (as of Jan 15, 2020). Gene annotations updated to Ensembl 99. See the [release notes](#) for details.

28-Feb-2020: We've added a new integration to the [NDEXProject](#) IQuery tool on our [Investigate Gene Sets](#) page. To use it from a Gene Set just click on the Advanced query 'Further investigate' link.

20-Aug-2019: MSigDB 7.0 released. This is a major release that includes a complete overhaul of gene symbol annotations, Reactome and GO gene sets, and corrections to miscellaneous errors. See the [release notes](#) for more information.

20-Aug-2019: GSEA 4.0.0 released. This release includes support for MSigDB 7.0, plus major internal updates for Java 11 support and performance improvements. See the [release notes](#) for more information.



License Terms

GSEA and MSigDB are available for use under [these license terms](#).

Please [register](#) to download the GSEA software, access our web tools, and view the MSigDB gene sets. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Citing GSEA

To cite your use of the GSEA software, a joint project of UC San Diego and Broad Institute, please reference Subramanian, Tamayo, et al. (2005, PNAS 102, 15545-15550) and Mootha, Lindgren, et al. (2003, Nat Genet 34, 267-273).

Funding

GSEA and MSigDB are currently funded by a grant from NCI's Informatics Technology for Cancer Research (ITCR).



GSEA

GSEA
Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact Team

<http://www.gsea-msigdb.org/gsea/index.jsp>

UC San Diego

BROAD
INSTITUTE

Overview

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether *a priori* defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes).

- ▶ [Download](#) the GSEA software and additional resources to analyze, annotate and interpret enrichment results.
- ▶ [Explore the Molecular Signatures Database \(MSigDB\)](#), a collection of annotated gene sets for use with GSEA software.
- ▶ [View documentation](#) describing GSEA and MSigDB.
- ▶ View guidelines for [using RNA-seq datasets with GSEA](#).
- ▶ Use the [GenePattern](#) platform to run analyses, including classical GSEA and a variation designed for single-sample analysis ([ssGSEA](#)).

What's New

19-Sep-2020: MSigDB 7.2 released. This release includes a substantial reorganization of C5 to accommodate the addition of the Human Phenotype Ontology, the addition of gene sets from WikiPathways to C2:CP, and the promotion of SCSig to C8, among other minor updates and additions. See the [release notes](#) for details.

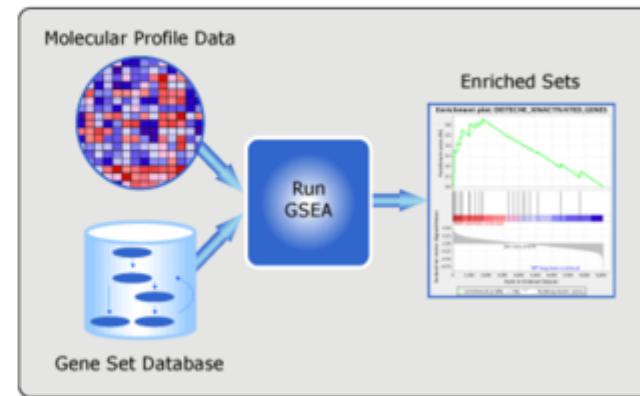
30-Jul-2020: GSEA 4.1.0 released. This release includes improved support for macOS Catalina, updated and improved Enrichment Reports, and numerous bug fixes. See the [release notes](#) for more information.

30-Mar-2020: MSigDB 7.1 released. New C3 sub-collections from the [Gene Transcription Regulation Database](#) and [miRDB](#). Updated gene sets from Reactome 72 and GO (as of Jan 15, 2020). Gene annotations updated to Ensembl 99. See the [release notes](#) for details.

28-Feb-2020: We've added a new integration to the [NDEXProject](#) IQuery tool on our [Investigate Gene Sets](#) page. To use it from a Gene Set just click on the Advanced query 'Further investigate' link.

20-Aug-2019: MSigDB 7.0 released. This is a major release that includes a complete overhaul of gene symbol annotations, Reactome and GO gene sets, and corrections to miscellaneous errors. See the [release notes](#) for more information.

20-Aug-2019: GSEA 4.0.0 released. This release includes support for MSigDB 7.0, plus major internal updates for Java 11 support and performance improvements. See the [release notes](#) for more information.



License Terms

GSEA and MSigDB are available for use under [these license terms](#).

Please [register](#) to download the GSEA software, access our web tools, and view the MSigDB gene sets. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Citing GSEA

To cite your use of the GSEA software, a joint project of UC San Diego and Broad Institute, please reference Subramanian, Tamayo, et al. (2005, PNAS 102, 15545-15550) and Mootha, Lindgren, et al. (2003, Nat Genet 34, 267-273).

Funding

GSEA and MSigDB are currently funded by a grant from NCI's Informatics Technology for Cancer Research (ITCR).



GSEA



Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact Team

Login to GSEA/MSigDB



Login

[Click here](#) to register to view the MSigDB gene sets and/or download the GSEA software. This helps us track and better serve our user community.

If you have already registered for GSEA or MSigDB please enter your registration email address below.

Items marked with * are required.

Email: *

GSEA



Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact Team

Login to GSEA/MSigDB



Login

[Click here](#) to register to view the MSigDB gene sets and/or download the GSEA software. This helps us track and better serve our user community.

If you have already registered for GSEA or MSigDB please enter your registration email address below.

Items marked with * are required.

Email: *

jon.sanchez@bsc.es



Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact Team

Login to GSEA/MSigDB



Login

[Click here](#) to register to view the MSigDB gene sets and/or download the GSEA software. This helps us track and better serve our user community.

If you have already registered for GSEA or MSigDB please enter your registration email address below.

Items marked with * are required.

Email: *

jon.sanchez@bsc.es

There are several options for GSEA software. All options implement exactly the same algorithm. Usage recommendations and installation instructions are listed below. For details on the GSEA algorithm and software refer to the [Documentation](#). For details on the latest release refer to the [Release Notes](#). The source is available from our [GitHub organization](#).

See the [license terms page](#) for details about the license for the GSEA software and source code. Please note that the license terms vary for different versions of the software.

GSEA v4.1.0 Mac App	Download and unzip the Mac App Archive then double-click the GSEA application to run it. You can move the app to the Applications folder or anywhere else.	download GSEA_4.1.0.app.zip
GSEA v4.1.0 for Windows	Download and run the installer. A GSEA shortcut will be created on the Desktop; double-click it to run the application. 64-bit Windows is required	download GSEA_Win_4.1.0-installer.exe
GSEA v4.1.0 for Linux	Download and unzip the Archive. See the included readme.txt for further instructions. 64-bit Linux is required	download GSEA_Linux_4.1.0.zip



GSEA Home Downloads Molecular Signatures Database Documentation Contact Team

Login to GSEA/MSigDB



Login

[Click here](#) to register to view the MSigDB gene sets and/or download the GSEA software. This helps us track and better serve our user community.

If you have already registered for GSEA or MSigDB please enter your registration email address below.

Items marked with * are required.

Email: *

jon.sanchez@bsc.es

There are several options for GSEA software. All options implement exactly the same algorithm. Usage recommendations and installation instructions are listed below. For details on the GSEA algorithm and software refer to the [Documentation](#). For details on the latest release refer to the [Release Notes](#). The source is available from our [GitHub organization](#).

See the [license terms page](#) for details about the license for the GSEA software and source code. Please note that the license terms vary for different versions of the software.

GSEA v4.1.0 Mac App	Download and unzip the Mac App Archive then double-click the GSEA application to run it. You can move the app to the Applications folder or anywhere else.	download GSEA_4.1.0.app.zip
GSEA v4.1.0 for Windows	Download and run the installer. A GSEA shortcut will be created on the Desktop; double-click it to run the application. 64-bit Windows is required	download GSEA_Win_4.1.0-installer.exe
GSEA v4.1.0 for Linux	Download and unzip the Archive. See the included readme.txt for further instructions. 64-bit Linux is required	download GSEA_Linux_4.1.0.zip

GSEA

GSEA 4.1.0 (Gene set enrichment analysis)

Home

Steps in GSEA analysis

- Load data
- Run GSEA
- Leading edge analysis
- Enrichment Map Visualization

Tools

- Run GSEAPreranked
- Collapse Dataset
- Chip2Chip mapping

Analysis history

GSEA reports

Processes: click 'status' field for results

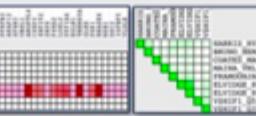
Name	Status
------	--------

Show results folder

5:05:24 PM | 4341 [INFO] - Made Vdb dir: /Users/jonsanchezvalle/gsea_home/output/feb15

50M of 1024M

Steps in GSEA

- What you need for GSEA
 - Expression data set
 - Phenotype annotation
 - Gene sets – use MSigDB or your own gene sets
- Run GSEA
 - Start with default parameters
 - If you want to collapse probes to genes, specify chip platform
- View results
- Leading edge analysis
 - Leading edge finds genes driving enrichment results

Gene Set Tools

- Chip2Chip mapping
 - Convert gene sets between platforms

Getting Help

GSEA web site:
www.gsea-msigdb.org

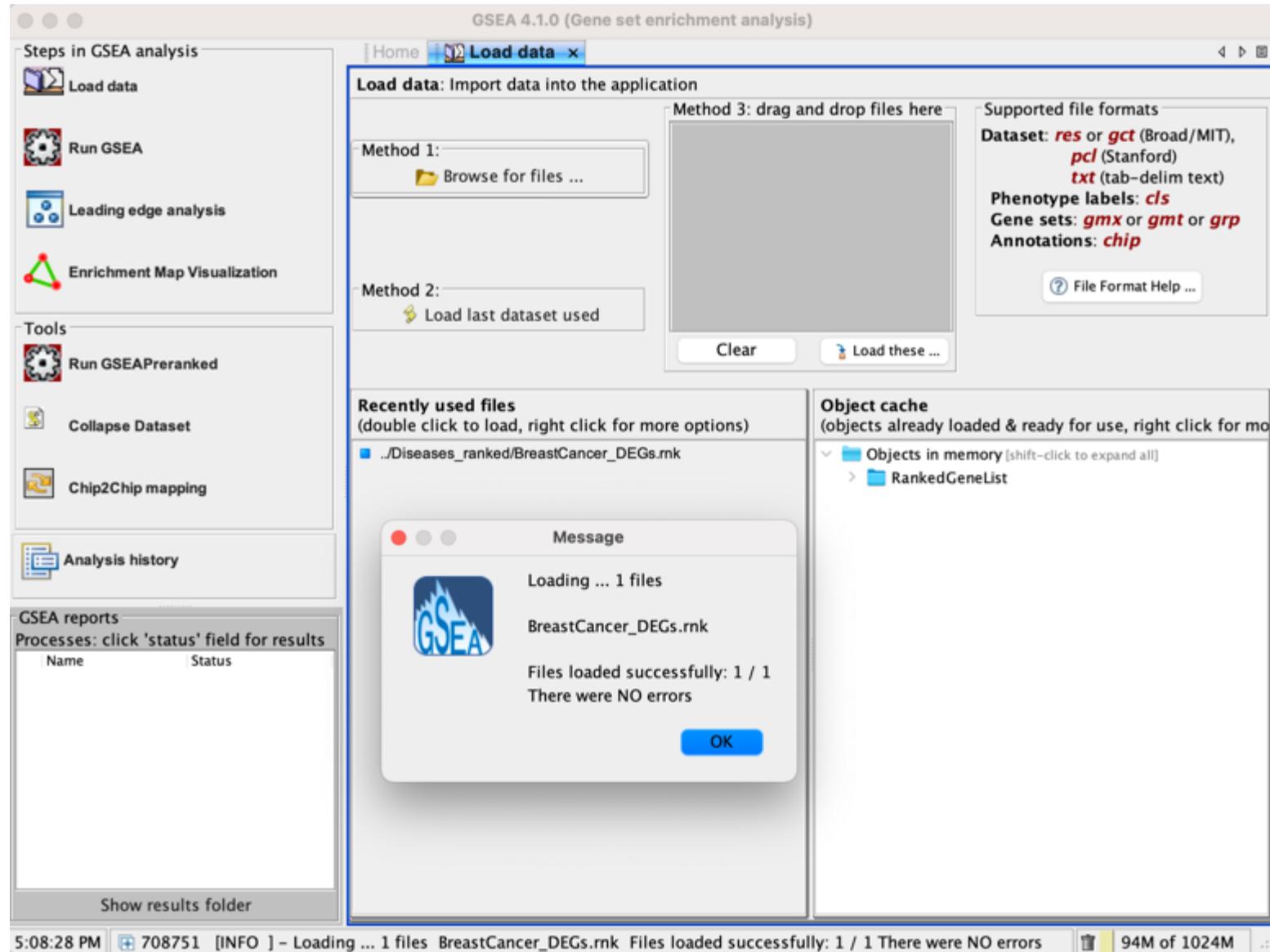
Contact the GSEA team:
gsea-msigdb.org/gsea/contact.jsp

BROAD INSTITUTE

UC San Diego

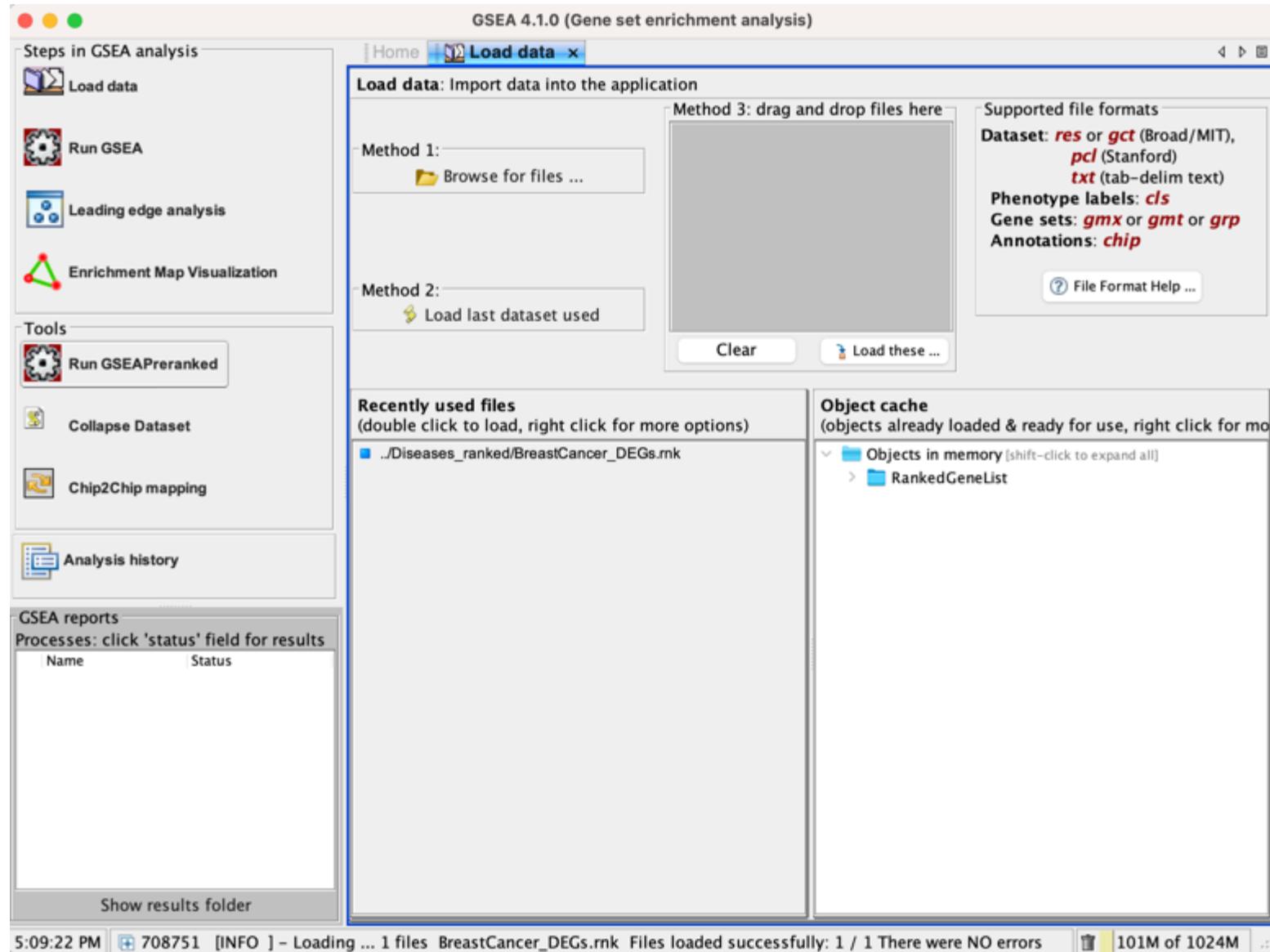
ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA



ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA



ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
⋮	
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA

GSEA 4.1.0 (Gene set enrichment analysis)

Home | Load data | Run Gsea on a Pre-Ranked gene list

GseaPreranked: Run GSEA on a pre-ranked (with external tools) gene list

Required fields

Gene sets database:

Number of permutations: 1000

Ranked List: BreastCancer_DEGs [13110 names]

Collapse/Remap to gene symbols: Remap_Only

Chip platform:

Basic fields

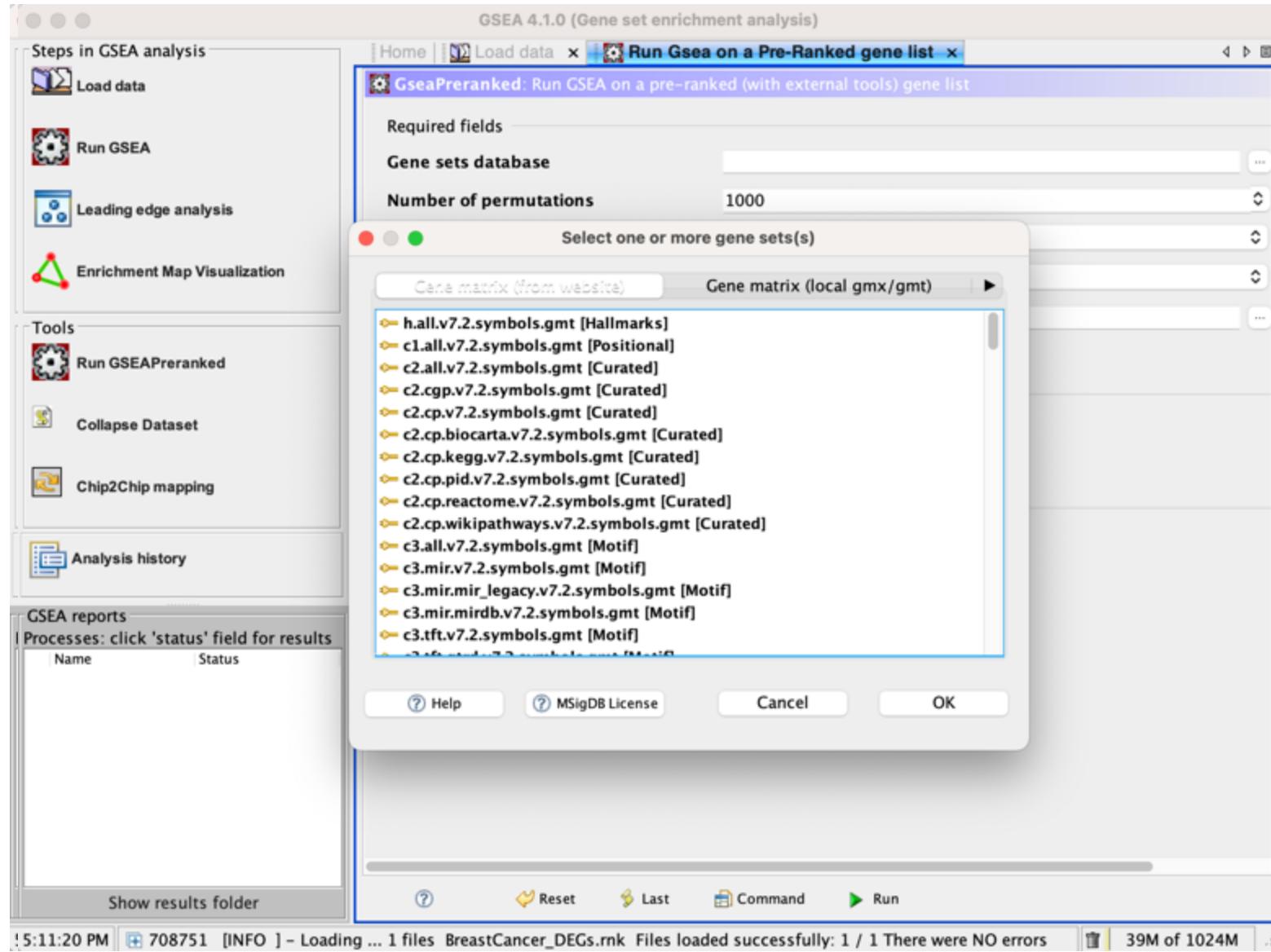
Advanced fields

Show results folder

5:09:56 PM | 708751 [INFO] - Loading ... 1 files BreastCancer_DEGs.rnk Files loaded successfully: 1 / 1 There were NO errors | 108M of 1024M

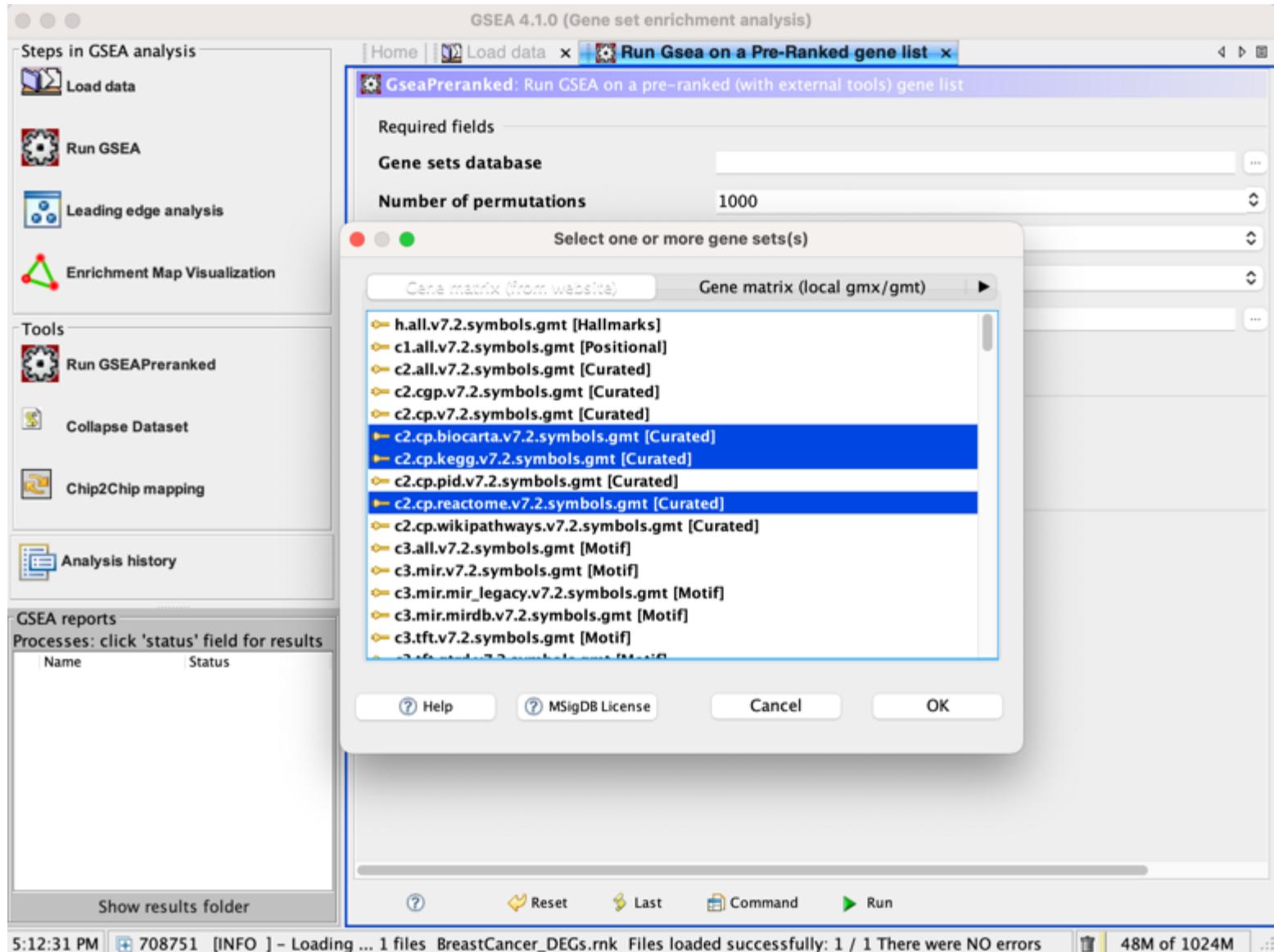
ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.76055557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA



ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA



ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA

GSEA 4.1.0 (Gene set enrichment analysis)

Home | Load data | Run Gsea on a Pre-Ranked gene list | GseaPreranked: Run GSEA on a pre-ranked (with external tools) gene list

Required fields

Gene sets database: dinstitute.org://pub/gsea/gene_sets/c5.go.mf.v7.2.symbols.gmt

Number of permutations: 1000

Ranked List: BreastCancer_DEGs [13110 names]

Collapse/Remap to gene symbols: Remap_Only

Chip platform:

Basic fields

Advanced fields

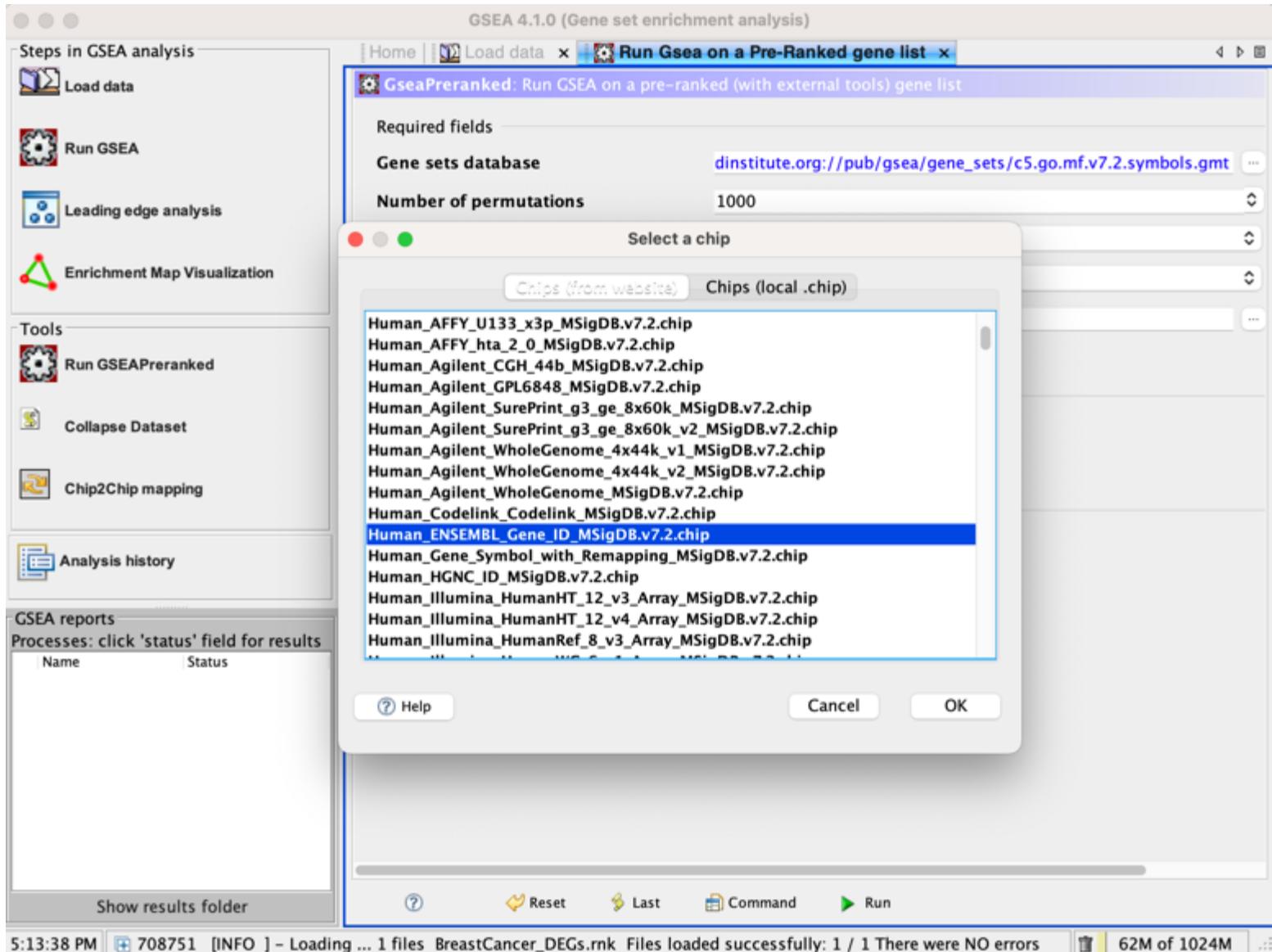
Show results folder

Reset Last Command Run

5:13:56 PM | 708751 [INFO] - Loading ... 1 files BreastCancer_DEGs.rnk Files loaded successfully: 1 / 1 There were NO errors | 64M of 1024M

ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.76055557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA



ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA

GSEA 4.1.0 (Gene set enrichment analysis)

Required fields

Gene sets database: dinstitute.org://pub/gsea/gene_sets/c5.go.mf.v7.2.symbols.gmt

Number of permutations: 1000

Ranked List: BreastCancer_DEGs [13110 names]

Collapse/Remap to gene symbols: Remap_Only

Chip platform: [rotations_versioned/Human_ENSEMBL_Gene_ID_MSigDB.v7.2.chip](#)

Basic fields

Analysis name: Breast_cancer_analysis

Enrichment statistic: weighted

Max size: exclude larger sets: 500

Min size: exclude smaller sets: 15

Save results in this folder: /Users/jonsanchezvalle/Desktop/GSEA_results

Advanced fields

Collapsing mode for probe sets => 1 gene: Max_probe

Normalization mode: meandiv

Alternate delimiter:

Create SVG plot images: false

Omit features with no symbol match: true

Show results folder

Reset Last Command Run

5:15:55 PM | 708751 [INFO] - Loading ... 1 files BreastCancer_DEGs.rnk Files loaded successfully: 1 / 1 There were NO errors | 140M of 1024M

ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA

GSEA 4.1.0 (Gene set enrichment analysis)

Steps in GSEA analysis

- Load data
- Run GSEA
- Leading edge analysis
- Enrichment Map Visualization

Tools

- Run GSEAPreranked
- Collapse Dataset
- Chip2Chip mapping

GSEA reports

Processes: click 'status' field for results

Name	Status
GseaPreranked	Running

Required fields

Gene sets database: dinstitute.org//pub/gsea/gene_sets/c5.go.mf.v7.2.symbols.gmt

Number of permutations: 1000

Ranked List: BreastCancer_DEGs [13110 names]

Collapse/Remap to gene symbols: Remap_Only

Chip platform: [rotations_versioned/Human_ENSEMBL_Gene_ID_MSigDB.v7.2.chip](#)

Basic fields

Progress...

Retrieving file Human_ENSEMBL_Gene_ID_MSigDB.v7.2.chip
58% complete

Cancel

Advanced fields

Collapsing mode for probe sets => 1 gene: Max_probe

Normalization mode: meandiv

Alternate delimiter:

Create SVG plot images: false

Omit features with no symbol match: true

Show results folder

Reset Last Command Run

5:16:37 PM | 1204725 [INFO] - File download started. Retrieving Human_ENSEMBL_Gene_ID_MSigDB.v7.2.chip from remote ser... | 168M of 1024M

ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA

Steps in GSEA analysis

- Load data
- Run GSEA
- Leading edge analysis
- Enrichment Map Visualization

Tools

- Run GSEAPreranked
- Collapse Dataset
- Chip2Chip mapping

GSEA reports

Processes: click 'status' field for results

Name	Status
1 GseaPreranked	Error!
2 GseaPreranked	Success

Show results folder

GSEA 4.1.0 (Gene set enrichment analysis)

Run Gsea on a Pre-Ranked gene list

Required fields

Gene sets database: dinstitute.org://pub/gsea/gene_sets/c5.go.mf.v7.2.symbols.gmt

Number of permutations: 1000

Ranked List: BreastCancer_DEGs [13110 names]

Collapse/Remap to gene symbols: Collapse

Chip platform: [rotations_versioned/Human_ENSEMBL_Gene_ID_MSigDB.v7.2.chip](#)

Basic fields

Analysis name: Breast_cancer_analysis

Enrichment statistic: weighted

Max size: exclude larger sets: 500

Min size: exclude smaller sets: 15

Save results in this folder: /Users/jonsanchezvalle/Desktop/GSEA_results

Advanced fields

Collapsing mode for probe sets => 1 gene: Max_probe

Normalization mode: meandiv

Alternate delimiter:

Create SVG plot images: false

Omit features with no symbol match: true

Reset Last Command Run

5:24:30 PM | 1659333 [INFO] - Made Vdb dir: /Users/jonsanchezvalle/gsea_home/reports_cache_April4_2006_build | 878M of 1279M

ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA

GSEA 4.1.0 (Gene set enrichment analysis)

Steps in GSEA analysis

- Load data
- Run GSEA
- Leading edge analysis
- Enrichment Map Visualization

Tools

- Run GSEAPreranked
- Collapse Dataset
- Chip2Chip mapping

Analysis history

GSEA reports

Processes: click 'status' field for results

Name	Status
1 GseaPreranked	Error!
2 GseaPreranked	Success

Required fields

Gene sets database: dinstitute.org://pub/gsea/gene_sets/c5.go.mf.v7.2.symbols.gmt

Number of permutations: 1000

Ranked List: BreastCancer_DEGs [13110 names]

Collapse/Remap to gene symbols: Collapse (circled in red)

Chip platform: [rotations_versioned/Human_ENSEMBL_Gene_ID_MSigDB.v7.2.chip](#)

Basic fields

Analysis name: Breast_cancer_analysis

Enrichment statistic: weighted

Max size: exclude larger sets: 500

Min size: exclude smaller sets: 15

Save results in this folder: /Users/jonsanchezvalle/Desktop/GSEA_results

Advanced fields

Collapsing mode for probe sets => 1 gene: Max_probe

Normalization mode: meandiv

Alternate delimiter:

Create SVG plot images: false

Omit features with no symbol match: true

Show results folder

Reset Last Command Run

5:24:30 PM | 1659333 [INFO] - Made Vdb dir: /Users/jonsanchezvalle/gsea_home/reports_cache_April4_2006_build | 878M of 1279M

ENSG00000131747	3.84114976014057
ENSG00000183856	3.64029042386646
ENSG00000088325	3.45214830097582
ENSG00000117724	3.39276680064398
ENSG00000105664	3.39183180008633
ENSG00000138778	2.98854055284515
ENSG00000167900	2.96068751211882
ENSG00000122952	2.8992268818348
ENSG00000166851	2.87074973136516
ENSG00000092853	2.78126437942206
ENSG00000258947	2.7605557221225
ENSG00000101412	2.75169237402904
ENSG00000135218	-3.44638638238693
ENSG00000181092	-3.46738437769294
ENSG00000101938	-3.5012015155528
ENSG00000165269	-3.50666133918497
ENSG00000168477	-3.54619290367672
ENSG00000187134	-3.6469134262705
ENSG00000167676	-3.67286108367659
ENSG00000151632	-3.69401899289397
ENSG00000170323	-3.75522012266886
ENSG00000173432	-3.79645835619139
ENSG00000166819	-3.81322197345373
ENSG00000162896	-3.87438345093075

GSEA

GSEA Report for Dataset BreastCancer_DEGs

Enrichment in phenotype: na

- 1412 / 5589 gene sets are upregulated in phenotype [na_pos](#)
- 738 gene sets are significant at FDR < 25%
- 463 gene sets are significantly enriched at nominal pvalue < 1%
- 617 gene sets are significantly enriched at nominal pvalue < 5%
- [Snapshot](#) of enrichment results
- Detailed [enrichment results in html](#) format
- Detailed [enrichment results in TSV](#) format (tab delimited text)
- [Guide to](#) interpret results

Enrichment in phenotype: na

- 4177 / 5589 gene sets are upregulated in phenotype [na_neg](#)
- 2747 gene sets are significantly enriched at FDR < 25%
- 1525 gene sets are significantly enriched at nominal pvalue < 1%
- 2082 gene sets are significantly enriched at nominal pvalue < 5%
- [Snapshot](#) of enrichment results
- Detailed [enrichment results in html](#) format
- Detailed [enrichment results in TSV](#) format (tab delimited text)
- [Guide to](#) interpret results

Dataset details

- The dataset has 13110 native features
- After collapsing features into gene symbols, there are: 12586 genes

Gene set details

- Gene set size filters (min=15, max=500) resulted in filtering out 6714 / 12303 gene sets
- The remaining 5589 gene sets were used in the analysis
- List of [gene sets used and their sizes](#) (restricted to features in the specified dataset)

Gene markers for the na_pos versus na_neg comparison

- The dataset has 12586 features (genes)
- Detailed [rank ordered gene list](#) for all features in the dataset

Global statistics and plots

- Plot of [p-values vs. NES](#)
- [Global ES histogram](#)

Other

- [Parameters](#) used for this analysis

GSEA

GSEA Report for Dataset BreastCancer_DEGs

Enrichment in phenotype: na

- 1412 / 5589 gene sets are upregulated in phenotype na_pos
- 738 gene sets are significant at FDR < 25%
- 463 gene sets are significantly enriched at nominal pvalue < 1%
- 617 gene sets are significantly enriched at nominal pvalue < 5%
- [Snapshot](#) of enrichment results
- Detailed [enrichment results in html](#) format
- Detailed [enrichment results in TSV](#) format (tab delimited text)
- [Guide to](#) interpret results

Enrichment in phenotype: na

- 4177 / 5589 gene sets are upregulated in phenotype na_neg
- 2747 gene sets are significantly enriched at FDR < 25%
- 1525 gene sets are significantly enriched at nominal pvalue < 1%
- 2082 gene sets are significantly enriched at nominal pvalue < 5%
- [Snapshot](#) of enrichment results
- Detailed [enrichment results in html](#) format
- Detailed [enrichment results in TSV](#) format (tab delimited text)
- [Guide to](#) interpret results

Dataset details

- The dataset has 13110 native features
- After collapsing features into gene symbols, there are: 12586 genes

Gene set details

- Gene set size filters (min=15, max=500) resulted in filtering out 6714 / 12303 gene sets
- The remaining 5589 gene sets were used in the analysis
- List of [gene sets used and their sizes](#) (restricted to features in the specified dataset)

Gene markers for the na_pos versus na_neg comparison

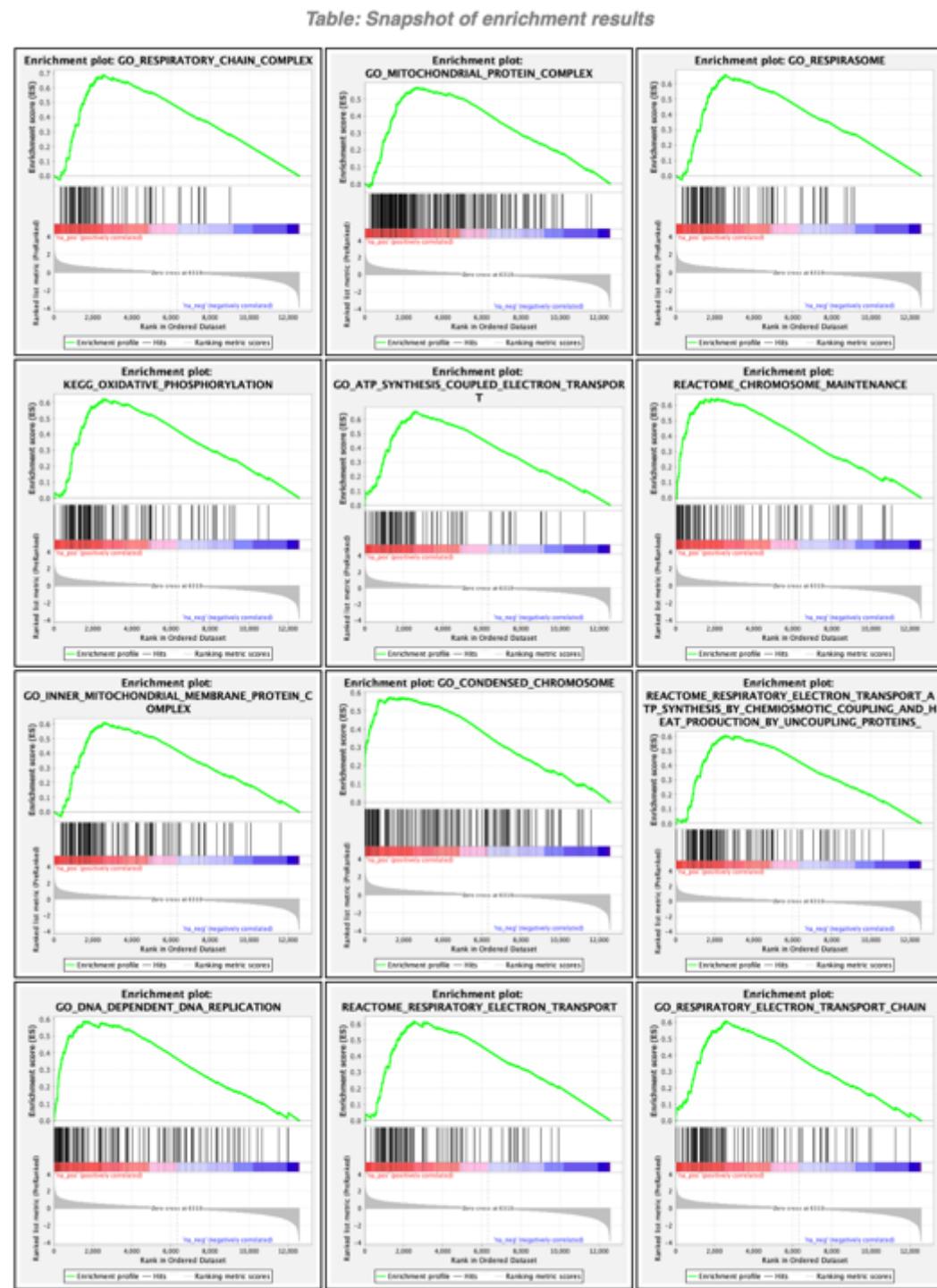
- The dataset has 12586 features (genes)
- Detailed [rank ordered gene list](#) for all features in the dataset

Global statistics and plots

- Plot of [p-values vs. NES](#)
- [Global ES histogram](#)

Other

- [Parameters](#) used for this analysis



GSEA

Table: GSEA Results Summary

Dataset	BreastCancer_DEGs_collapsed
Phenotype	NoPhenotypeAvailable
Upregulated in class	na_pos
GeneSet	KEGG_OXIDATIVE_PHOSPHORYLATION
Enrichment Score (ES)	0.62362283
Normalized Enrichment Score (NES)	2.8484964
Nominal p-value	0.0
FDR q-value	0.0

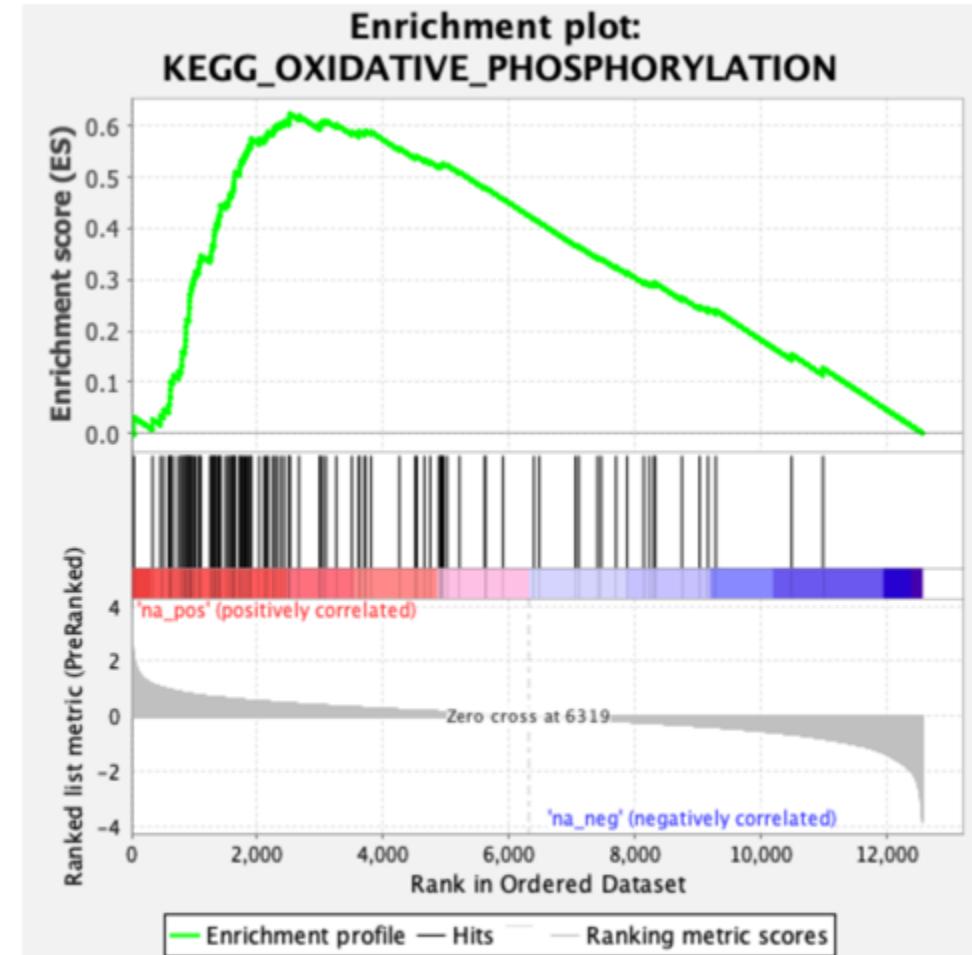


Fig 1: Enrichment plot: KEGG_OXIDATIVE_PHOSPHORYLATION
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

GSEA

	Symbol	Title	Rank in Gene List	Rank Metric Score	Running ES	Core Enrichment
1	COX6C	cytochrome c oxidase subunit 6C [Source:HGNC Symbol;Acc:HGNC:2285]	42	2.086	0.0310	Yes
2	MT-CO2	mitochondrially encoded cytochrome c oxidase II [Source:HGNC Symbol;Acc:HGNC:7421]	337	1.154	0.0264	Yes
3	ATP6V0B	ATPase H ⁺ transporting V0 subunit b [Source:HGNC Symbol;Acc:HGNC:861]	460	1.054	0.0340	Yes
4	NDUFA7	NADH:ubiquinone oxidoreductase subunit A7 [Source:HGNC Symbol;Acc:HGNC:7691]	514	1.014	0.0464	Yes
5	NDUFB1	NADH:ubiquinone oxidoreductase subunit B1 [Source:HGNC Symbol;Acc:HGNC:7695]	595	0.969	0.0560	Yes
6	ATP5MF	ATP synthase membrane subunit f [Source:HGNC Symbol;Acc:HGNC:848]	611	0.960	0.0706	Yes
7	COX6B1	cytochrome c oxidase subunit 6B1 [Source:HGNC Symbol;Acc:HGNC:2280]	622	0.953	0.0854	Yes
8	NDUFS8	NADH:ubiquinone oxidoreductase core subunit S8 [Source:HGNC Symbol;Acc:HGNC:7715]	626	0.952	0.1009	Yes
9	UQCRQ	ubiquinol-cytochrome c reductase complex III subunit VII [Source:HGNC Symbol;Acc:HGNC:29594]	669	0.926	0.1127	Yes
10	ATP6V0E2	ATPase H ⁺ transporting V0 subunit e2 [Source:HGNC Symbol;Acc:HGNC:21723]	749	0.889	0.1211	Yes
11	NDUFB4	NADH:ubiquinone oxidoreductase subunit B4 [Source:HGNC Symbol;Acc:HGNC:7699]	793	0.862	0.1318	Yes
12	NDUFS6	NADH:ubiquinone oxidoreductase subunit S6 [Source:HGNC Symbol;Acc:HGNC:7713]	805	0.853	0.1450	Yes
13	NDUFB9	NADH:ubiquinone oxidoreductase subunit B9 [Source:HGNC Symbol;Acc:HGNC:7704]	808	0.852	0.1588	Yes
14	NDUFS5	NADH:ubiquinone oxidoreductase subunit S5 [Source:HGNC Symbol;Acc:HGNC:7712]	852	0.833	0.1691	Yes
15	ATP6AP1	ATPase H ⁺ transporting accessory protein 1 [Source:HGNC Symbol;Acc:HGNC:868]	860	0.828	0.1822	Yes
16	ATP5ME	ATP synthase membrane subunit e [Source:HGNC Symbol;Acc:HGNC:846]	867	0.826	0.1953	Yes
17	NDUFC2	NADH:ubiquinone oxidoreductase subunit C2 [Source:HGNC Symbol;Acc:HGNC:7706]	870	0.825	0.2087	Yes
18	UQCR10	ubiquinol-cytochrome c reductase, complex III subunit X [Source:HGNC Symbol;Acc:HGNC:30863]	877	0.821	0.2218	Yes
19	ATP6V1F	ATPase H ⁺ transporting V1 subunit F [Source:HGNC Symbol;Acc:HGNC:16832]	916	0.806	0.2320	Yes
20	ATP5MC1	ATP synthase membrane subunit c locus 1 [Source:HGNC Symbol;Acc:HGNC:841]	918	0.806	0.2452	Yes

GSEA

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val
1	GO_RESPIRATORY_CHAIN_COMPLEX	Details ...	74	0.69	2.95	0.000	0.000
2	GO_MITOCHONDRIAL_PROTEIN_COMPLEX	Details ...	253	0.57	2.92	0.000	0.000
3	GO_RESPIRASOME	Details ...	88	0.66	2.88	0.000	0.000
4	KEGG_OXIDATIVE_PHOSPHORYLATION	Details ...	114	0.62	2.85	0.000	0.000
5	GO_ATP_SYNTHESIS_COUPLED_ELECTRON_TRANSPORT	Details ...	88	0.65	2.81	0.000	0.000
6	REACTOME_CHROMOSOME_MAINTENANCE	Details ...	89	0.64	2.75	0.000	0.000
7	GO_INNER_MITOCHONDRIAL_MEMBRANE_PROTEIN_COMPLEX	Details ...	128	0.61	2.75	0.000	0.000
8	GO_CONDENSED_CHROMOSOME	Details ...	156	0.57	2.74	0.000	0.000
9	REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS	Details ...	119	0.61	2.73	0.000	0.000
10	GO_DNA_DEPENDENT_DNA_REPLICATION	Details ...	129	0.59	2.72	0.000	0.000

```
for a in `ls Ranked/`; do java -Xmx2000m -cp ./gsea.jar xtools.gsea.GseaPreranked -gmx gseaftp.broadinstitute.org://pub/gsea/gene_sets_final/c2.cp.kegg.v7.0.symbols.gmt,gseaftp.broadinstitute.org://pub/gsea/gene_sets_final/c2.cp.reactome.v7.0.symbols.gmt,gseaftp.broadinstitute.org://pub/gsea/gene_sets_final/c5.bp.v7.0.symbols.gmt -collapse true -mode Max_probe -norm meandiv -nperm 1000 -rnk /Users/jonsanchezvalle/Desktop/Pathways/Ranked/$a -scoring_scheme weighted -rpt_label ${a}/%?????} -chip gseaftp.broadinstitute.org://pub/gsea/annotations/ENSEMBL_human_gene.chip -include_only_symbols true -make_sets true -plot_top_x 20 -rnd_seed timestamp -set_max 500 -set_min 15 -zip_report false -out /Users/jonsanchezvalle/Desktop/Pathways/Results -gui false; done
```

GSEA

GCT: Gene Cluster Text file format (*.gct)

The GCT format is a tab delimited file format that describes an expression dataset. It is organized as follows:

of samples

Third column onward
are sample names
These must be UNIQUE

Microsoft Excel - ExampleA_dataset_HG_U133A

E15 fx 102.9

A	B	C	D	E	F	G
1 #1.2						
2 1000	130					
3 NAME	Description	DLBCL 205	DLBCL 206	DLBCL 232	DLBCL 239	DLBCL 240
4 1007_s_at	U48705 /FEATURE=mRNA /DEFINITION=HS	280.53	271.48	113.57	124.91	124.91
5 1053_at	M87338 /FEATURE= /DEFINITION=HUMA1S	32.13	91.6	117.43	41.29	33.66
6 117_at	X51757 /FEATURE=cds /DEFINITION=HSPT0	51.27	61.12	24.1	41.44	43.56
7 121_at	X69699 /FEATURE= /DEFINITION=HSPAX8A	738.32	330.59	249.89	394.55	329.55
8 1255_g_at	L36861 /FEATURE=expanded_cds /DEFINITION=	88.45	12.94	18.46	29.96	39
9 1294_at	L13852 /FEATURE= /DEFINITION=HUME1UR	85.57	88.06	62.24	96.59	81.01
10 1316_at	X55005 /FEATURE=mRNA /DEFINITION=HS	106.87	45.11	30.05	46.65	36.5
11 1320_at	X79510 /FEATURE=cds /DEFINITION=HSPTF	58.49	27.95	17.6	27.87	26.52
12 1405_i_at	M21121 /FEATURE= /DEFINITION=HUMTCSI	10.83	135.24	13.43	203.16	85.74
13 1431_at	J02843 /FEATURE=cds /DEFINITION=HUMC	41.88	24.09	16.07	26.68	25.4
14 1438_at	X75208 /FEATURE=cds /DEFINITION=HSPTK	80.87	9.77	15.33	11.18	44.59
15 1487_at	L38487 /FEATURE=mRNA /DEFINITION=HU	64.26	80.61	102.9	59.77	105.72
16 1494_f_at	M33318 /FEATURE=mRNA /DEFINITION=HU	213.37	96.88	65.06	96.14	78.77
17 1598_g_at	L13720 /FEATURE= /DEFINITION=HUMGAS	458.88	215.59	186.72	187.36	237.69
18 160020_at	Z48481 /FEATURE=cds /DEFINITION=HSMM	411.94	171.16	130	234.76	266.96
19 1729_at	L41690 /FEATURE= /DEFINITION=HUMTRAD	81.59	83.94	74.75	110.9	126.98
20 1773_at	L00635 /FEATURE= /DEFINITION=HUMPTPE	62.82	45.96	41.15	23.1	28.41
21 177_at	U38545 /FEATURE= /DEFINITION=HSU3854	57.04	28.05	16.74	29.66	53.29
22 179_at	U38980 /FEATURE= /DEFINITION=U38980_H	333.96	254.15	241.24	350.58	13.53
23 1804_at	U38980 /FEATURE= /DEFINITION=U38980_H	99.241	99.491	12.661	99.841	16.961

If editing, in Excel, make sure to save your data as “tab delimited text”

CLS: Categorical (e.g tumor vs normal) class file format (*.cls)

Always 1

GMX: Gene MatriX file format (*.gmx)

First row are gene set names. Duplicates are not allowed

Second row contains a brief description. Its optional – you can fill in a dummy field (e.g. "na")

Each column represents one gene set

Unequal lengths (i.e. # of genes) is allowed

	RECOL4						
	A	B	C	D	E	F	G
1	chr10q24	chr5q23	chr8q24	chr16q24	chr13q14	chr7p21	chr10q23
2	na	na	na	na	na	na	na
3	PITX3	ALDH7A1	HAS2	RPL13	AKAP11	ARL4A	SNCG
4	SPFH1	IL13	LRRK14	GALNS	ARL11	SCIN	FER1L3
5	NEURL	8-Sep		TSTA3	FANCA	ATP7B	GLCC1
6	C10orf12	IRF1	DGAT1	CPNE7	C13orf1	SP8	HHEX
7	NDUFB8	ACSL6	RECOL4		C13orf9	SOSTDC1	TNKS2
8	C10orf95	IL4	GPR172A		CAB39L	TM4SF13	MPHOSPH1
9	DNTT	SLC12A2	COL14A1		CDADC1	FERD3L	CYP2C18
10	USMG5	PPIC	EXT1		CHC1L	ANKMY2	C10orf117
11	CWF19L1	CSF2	RAD21		CKAP2	ICA1	MINPP1
12	SUFU	SLC22A5	SLA		COG3	TWIST1	LRRC21
13	OBFC1	CSNK1G3			COG6	DGKB	PDLIM1
14	PEO1	DMXL1			CPB2	NDUFA4	HELLS
15	PIK3AP1	P4HA2			CYSLTR2		CH25H
16	UBTD1	ZNF608			DDX26		LDB3
17	CUTC	LOX			DES		PPR30
18	SEC31L2	FTMT			DGKH		
19	AS3MT	ADAMTS19			DLEU1		
20	MGEA5	IL5			DNAJ1		
21	NPM3	IL3					

GSEA (*available metrics*)

For **categorical phenotypes**, GSEA determines a gene's mean expression value for each phenotype and then uses one of the following metrics to calculate the gene's differential expression with respect to the two phenotypes. To use median rather than mean expression values, set the *Median for class metrics* parameter to True, as described above.

- Signal2Noise (default) uses the difference of means scaled by the standard deviation. **Note:** You must have at least three samples for each phenotype to use this metric.

$$\frac{\mu_A - \mu_B}{\sigma_A + \sigma_B}$$

where μ is the mean and σ is the standard deviation; σ has a minimum value of $.2 * \text{absolute}(\mu)$, where $\mu=0$ is adjusted to $\mu=1$. The larger the signal-to-noise ratio, the larger the differences of the means (scaled by the standard deviations); that is, the more distinct the gene expression is in each phenotype and the more the gene acts as a "class marker."

- tTest uses the difference of means scaled by the standard deviation and number of samples. **Note:** You must have at least three samples for each phenotype to use this metric.

$$\frac{\mu_A - \mu_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$$

where μ is the mean, n is the number of samples, and σ is the standard deviation; σ has a minimum value of $.2 * \text{absolute}(\mu)$, where $\mu=0$ is adjusted to $\mu=1$. The larger the tTest ratio, the more distinct the gene expression is in each phenotype and the more the gene acts as a "class marker."

- Ratio_of_Classes (also referred to as fold change) uses the ratio of class means to calculate fold change for natural scale data:

$$\frac{\mu_A}{\mu_B}$$

where μ is the mean. The larger the fold change, the more distinct the gene expression is in each phenotype and the more the gene acts as a "class marker."

- Diff_of_Classes uses the difference of class means to calculate fold change for log scale data:

$$\mu_A - \mu_B$$

where μ is the mean. The larger the fold change, the more distinct the gene expression is in each phenotype and the more the gene acts as a "class marker."

- log2_Ratio_of_Classes uses the log2 ratio of class means to calculate fold change for natural scale data:

$$\log_2\left(\frac{\mu_A}{\mu_B}\right)$$

where μ is the mean. This is the recommended statistic for calculating fold change for natural scale data.

Other methods for gene set enrichments

FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes FREE

[Fátima Al-Shahrour](#) , [Ramón Díaz-Uriarte](#), [Joaquín Dopazo](#)

ROMA: Representation and Quantification of Module Activity from Target Expression Data

[Loredana Martignetti](#),^{1,2,3,4} [Laurence Calzone](#),^{1,2,3,4} [Eric Bonnet](#),^{1,2,3,4} [Emmanuel Barillot](#),^{1,2,3,4} and
[Andrei Zinovyev](#)^{1,2,3,4,*}

Visualization

OPEN ACCESS Freely available online

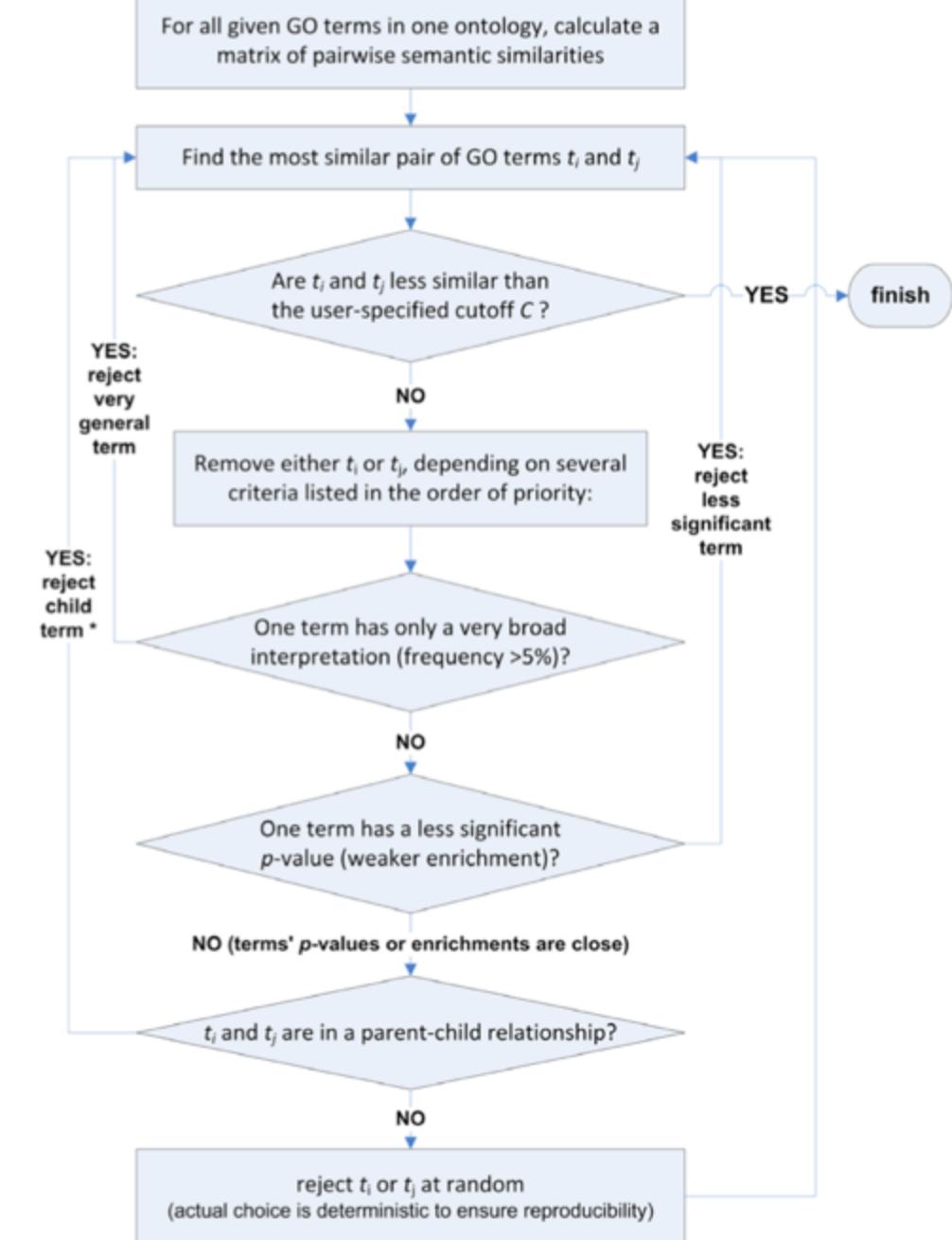
PLOS one

REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms

Fran Supek^{1,2*}, Matko Bošnjak¹, Nives Škunca¹, Tomislav Šmuc¹



term ID	description	frequency	pini	log10 p-value	uniqueness	dispensability
GO:7049	cell cycle	3.652 %	-14.2652	0.91	0.00	
GO:30949	positive regulation of vascular endothelial growth factor receptor signaling pathway	0.036 %	-3.3958	0.85	0.02	
GO:51656	establishment of organelle localization	0.260 %	-3.8570	0.85	0.02	
GO:31239	establishment of spindle localization	0.045 %	-3.3375	0.60	0.94	
GO:49000	establishment of mitotic spindle localization	0.077 %	-4.0070	0.41	0.77	
GO:7059	chromosome segregation	0.287 %	-10.9872	0.92	0.03	



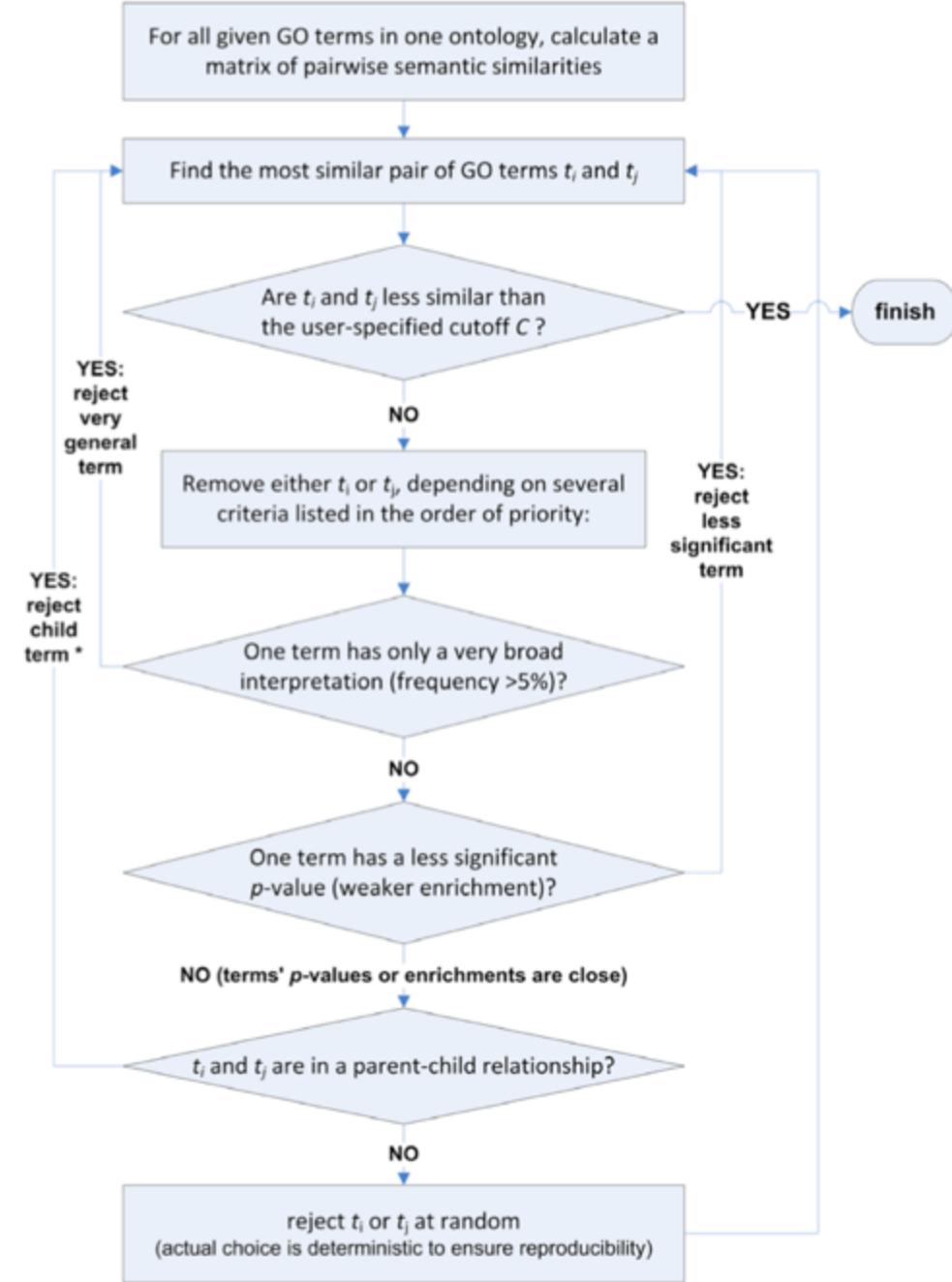
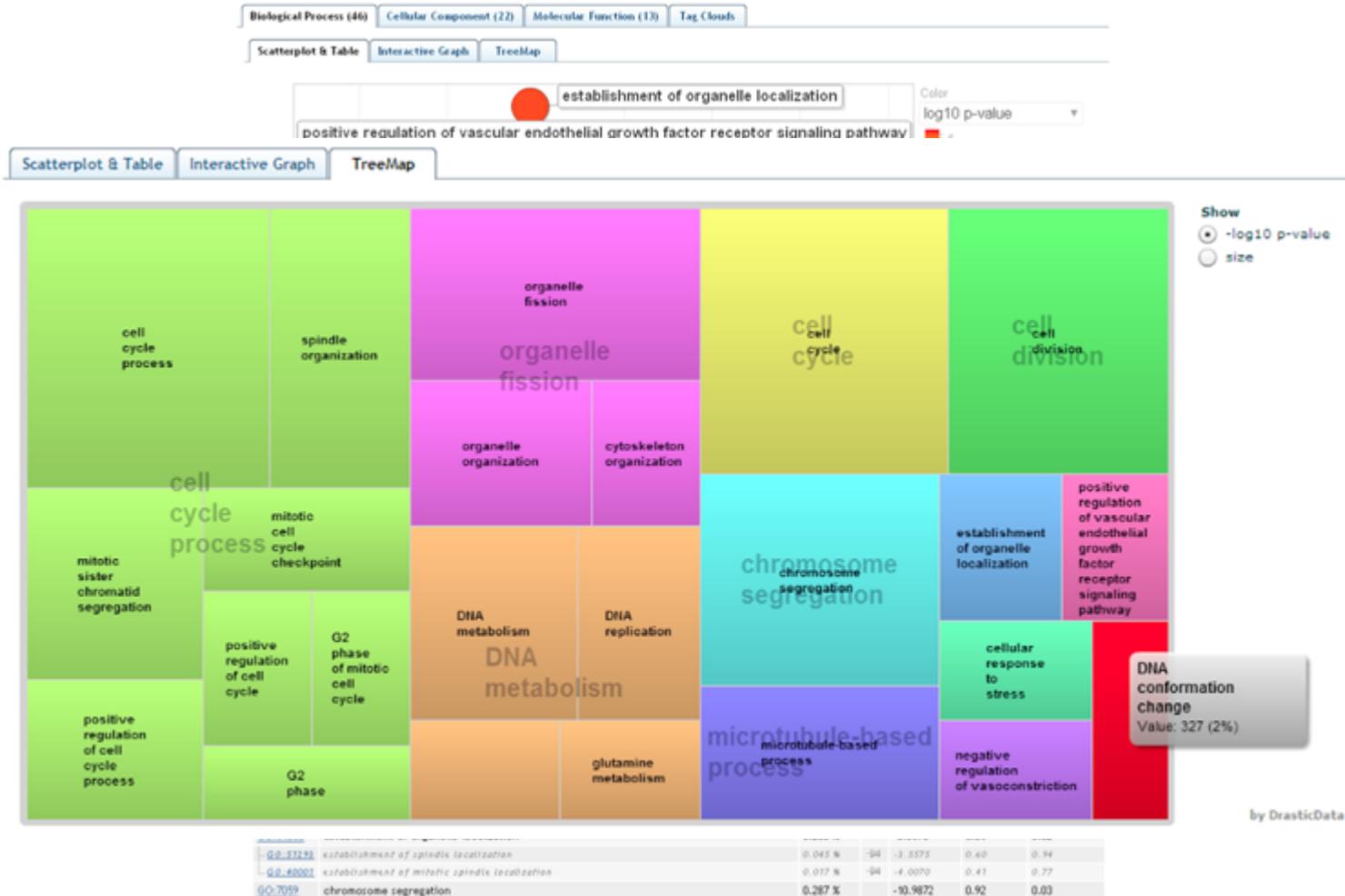
Visualization

OPEN  ACCESS Freely available online

PLOS one

REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms

Fran Supek^{1,2*}, Matko Bošnjak¹, Nives Škunca¹, Tomislav Šmuc¹



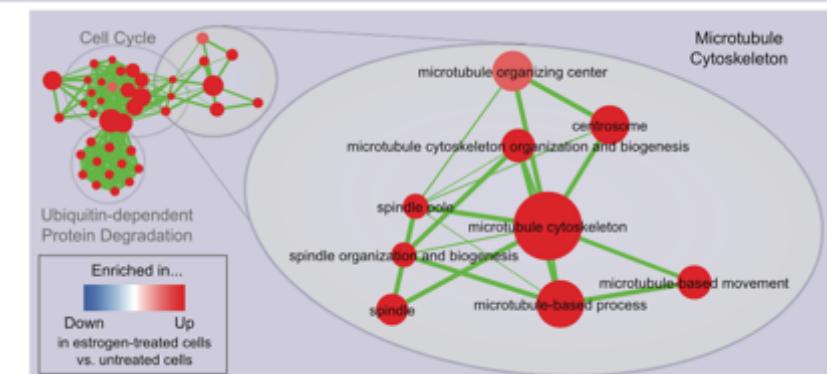
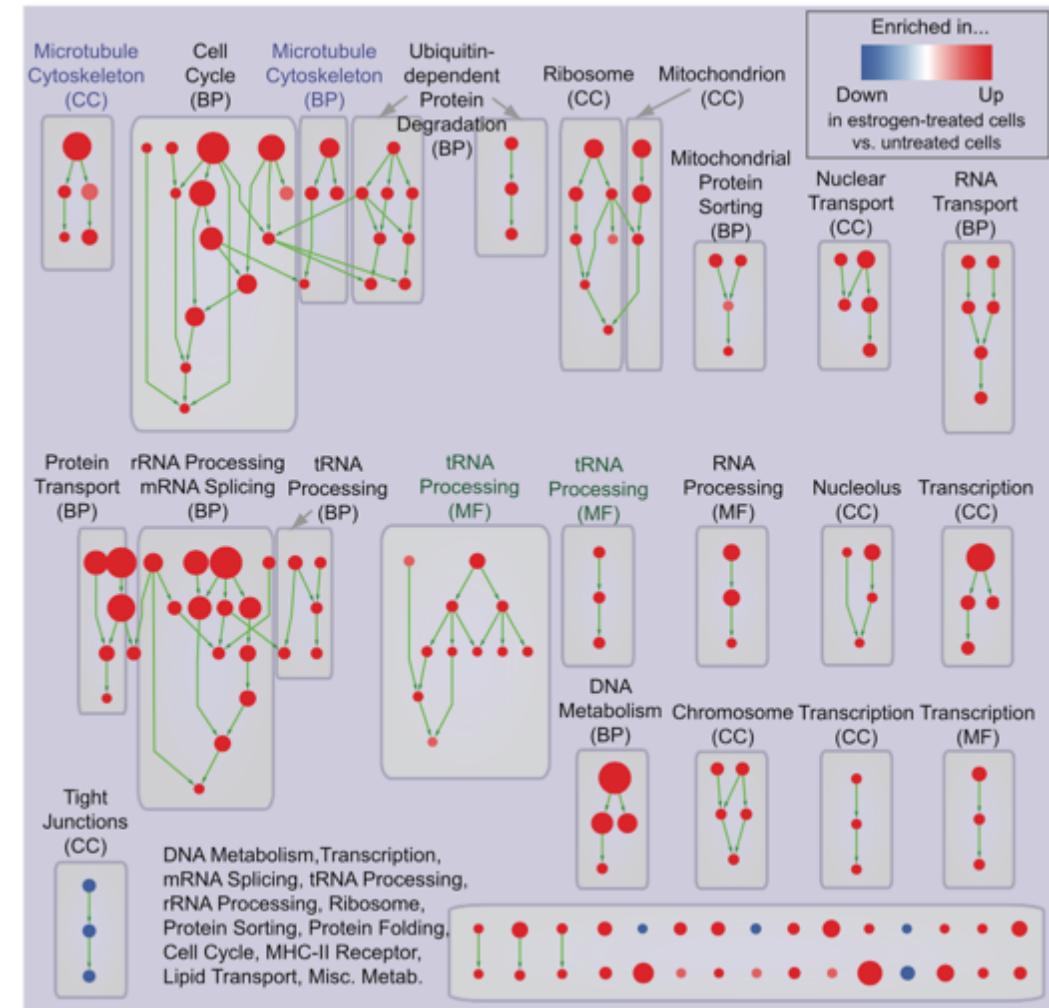
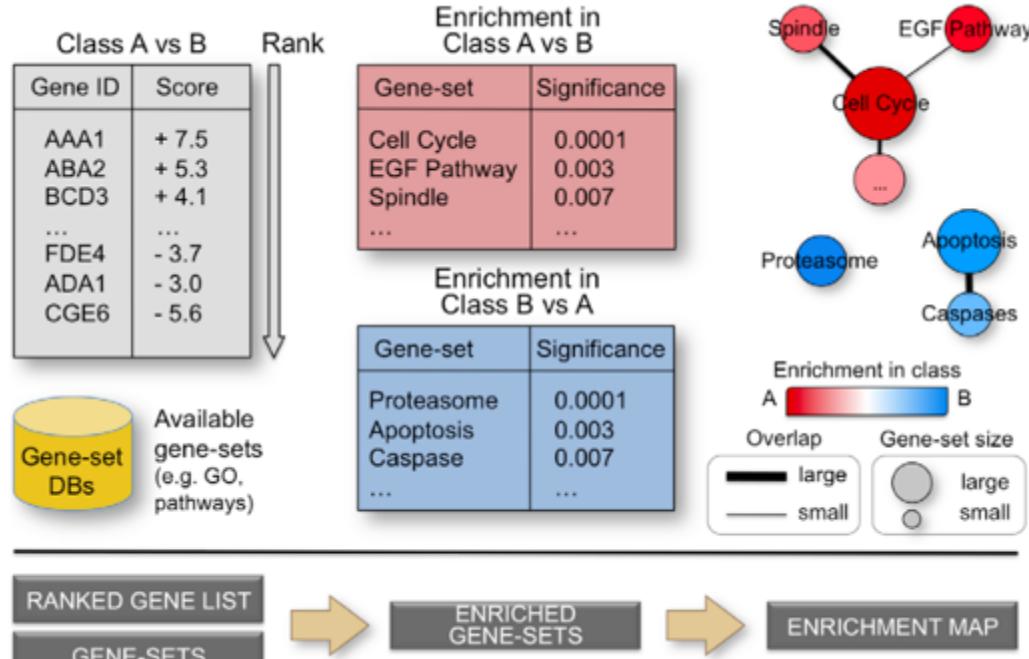
Visualization

OPEN ACCESS Freely available online



Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation

Daniele Merico*, Ruth Isserlin, Oliver Stueker, Andrew Emili, Gary D. Bader*



Other methods

A comparison of mechanistic signaling pathway activity analysis methods

Alicia Amadoz, Marta R. Hidalgo, Cankut Çubuk, José Carbonell-Caballero and Joaquín Dopazo

