

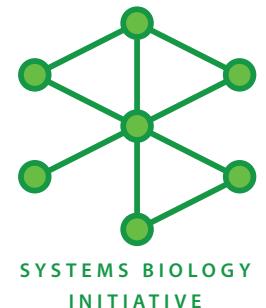
BIOPLATFORMS
AUSTRALIA



Biological Interpretation

Presenter | Susan Corley, SBI UNSW

Monash University, Melbourne
15th June 2016



EMBL-EBI



Gene ontology and network analysis

There are many tools available to perform gene ontology analysis.

To demonstrate the fundamental approaches we will use:

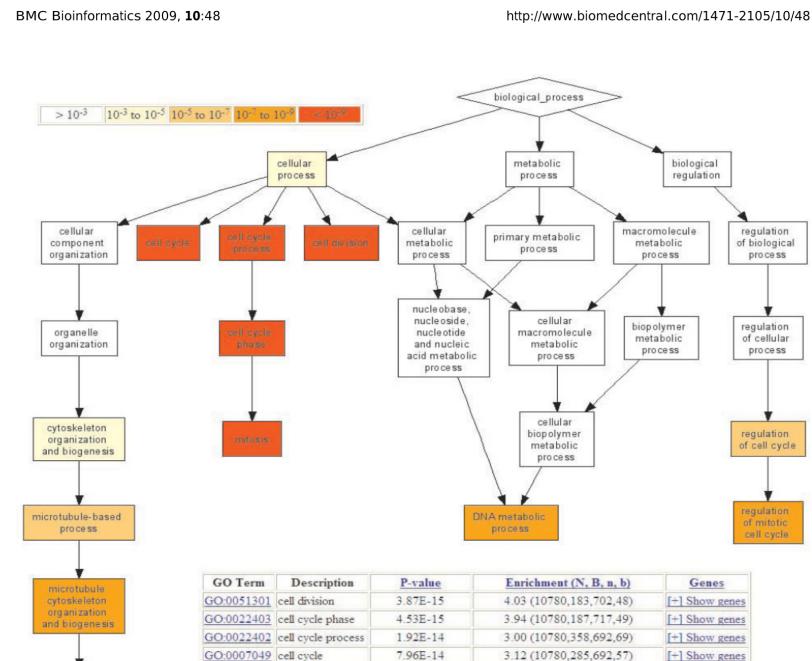
- the Goana function within limma (Bioconductor package)
- The web tool DAVID <https://david.ncifcrf.gov>
- The web tool Gorilla <http://cbl-gorilla.cs.technion.ac.il>
- REVIGO <http://revigo.irb.hr>
- STRING <http://string-db.org> (web tool and Bioconductor package)

What is gene ontology?

- GO annotations: are statements describing the functions of specific genes.
- The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway.
- Each statement is based on a specified piece of evidence.

- Three categories Biological Process (BP), Molecular Function (MF), Cellular Component (CC)
- GO terms are arranged in a hierarchical manner

(see <http://geneontology.org>)



Goana

- Function available within the limma package

- `DE_GOana<-goana(fit_v, coef=2, geneid=fit_v$genes$Entrez, FDR=0.05, species = "Hs", trend=F, plot=F)`
- `DE_GOana_top_BP_down<- topGO(DE_GOana, ontology=c("BP"), sort = "down", number=150L, truncate.term=50)`
- `head(DE_GOana_top_BP_down, 10)`

	Term	Ont	N	Up	Down	P.Up	P.Down
GO:0044763	single-organism cellular process	BP	8472	1359	1160	3.9E-02	7.7E-14
GO:0044699	single-organism process	BP	9303	1483	1238	6.1E-02	6.4E-11
GO:0007049	cell cycle	BP	1453	249	255	5.2E-02	1.6E-10
GO:0000278	mitotic cell cycle	BP	899	151	172	1.7E-01	3.8E-10
GO:0022402	cell cycle process	BP	1120	183	200	2.6E-01	5.2E-09
GO:0006613	cotranslational protein targeting to membrane	BP	104	18	34	3.6E-01	3.2E-08
GO:0051301	cell division	BP	540	78	109	8.0E-01	4.6E-08
GO:0006614	SRP-dependent cotranslational protein targeting...	BP	102	17	33	4.3E-01	6.8E-08
GO:0016043	cellular component organization	BP	4676	749	664	1.9E-01	1.0E-07
GO:1903047	mitotic cell cycle process	BP	760	132	141	9.7E-02	1.2E-07

Goana

- Function available within the limma package
 - `DE_GOana_top_BP_up<- topGO(DE_GOana, ontology=c("BP"), sort = "up", number=150L, truncate.term=50)`
 - `head(DE_GOana_top_BP_up, 10)`
 - `DE_GOana_top_BP<- topGO(DE_GOana, ontology=c("BP"), number=150L, truncate.term=50)`
 - `head(DE_GOana_top_BP, 10)`
- Change “BP” (biological process) to “MF” (molecular function) and “CC” (cellular component)

Web based tools

Data set to be examined

Open voom_res_sig_lfc.txt with LibreOffice Calc

The screenshot shows the LibreOffice Calc interface. On the left, the 'Text Import - [voom_res_sig_lfc.txt]' dialog box is open, displaying settings for importing the file. The 'Character set' is set to 'Unicode (UTF-8)', 'Language' to 'Default - English (USA)', and 'From row' to '1'. Under 'Separator options', 'Separated by' is selected, with 'Tab' checked. Other options like 'Semicolon' and 'Space' are unchecked. The 'Text delimiter' is set to '\"'. In the 'Other options' section, 'Quoted field as text' is checked, while 'Detect special numbers' is unchecked. The 'Column type' dropdown is set to 'Standard'. Below the dialog, a preview of the data is shown in a table.

The right side of the screen shows the LibreOffice desktop environment. A file browser window titled 'edgeR' is open, showing files in the 'Computer' section. The file 'voom_res_sig_lfc.txt' is highlighted with a yellow border. The status bar at the bottom of the desktop window indicates that 'voom_res_sig_lfc.txt' was selected (121.7 kB).

	Standard	Standard	Standard	Standard	Standard
1	logFC	AveExpr	t	P.Value	
2	ENSG00000096060	4.9996118424	8.2501016823	49.307330098	1.98559
3	ENSG00000151503	5.8186917098	7.5790957582	46.9562397078	3.59564
4	ENSG00000115648	2.6217227436	10.921999583	43.754712478	8.47791
5	ENSG00000064042	1.6681132578	11.1664100218	39.2933687672	3.12595
6	ENSG00000162772	3.3253151411	8.884256552	37.7029038816	5.15801
7	ENSG00000140526	2.2372479612	9.9303559625	37.541490348	5.43324
8	ENSG00000156451	1.6532052607	7.2967567401	35.2162467202	1.17040

Copy the Ensembl column to be pasted into the tools we use

A	B	C	D	E	F	G
1	logFC	AveExpr	t	P_Value	adj_P_Val	B
2	4.9996118424	8.2501016823	49.307330098	1.98559728741774e-15	2.965325554523e-11	24.6419070252
3	5.8186917098	7.5790957582	46.9562397078	3.5956415114866e-15	2.965325554523e-11	23.3518548348
4	2.6217227436	10.921999583	43.754712478	8.47791054119655e-15	4.66115521554986e-11	24.5385743916
5	1.6681132578	11.1664100218	39.2933687672	3.12595886083593e-14	1.2889891362657e-10	23.2203916905
6	3.3253151411	8.8842565552	37.7029038816	5.15801853229298e-14	1.49359927951086e-10	22.5239749633
7	2.2372479612	9.9303559625	37.541490348	5.43324583306971e-14	1.49359927951086e-10	22.6685730391
8	4.6520052607	7.3267567401	35.2162467262	1.17848308063179e-13	2.77684284742011e-10	21.0129737221
9	2.5840317218	9.4310722903	34.5629991329	1.478163676661e-13	3.04760396035624e-10	21.6518004207
10	3.256417992	7.9670154305	33.6178203194	2.067161250506e-13	3.78841751842733e-10	21.0177452347
11	2.3140301099	9.4816076674	31.2324874152	5.03013679042374e-13	7.54246147465902e-10	20.4614426502
12	2.1144218261	9.2186231829	30.3953464877	6.98218776540902e-13	9.5970170835547e-10	20.1287624068
13	2.8156833561	7.5596398005	28.8170628559	1.32803087057941e-12	1.6849647061028e-09	19.2626071536
14	-2.3073873963	10.9581346828	-28.25852265	1.68112644357335e-12	1.98060711144992e-09	19.2135671005
15	1.7749317324	9.9072152413	27.9012006236	1.95948586416919e-12	2.15465065624044e-09	19.0874008433
16	2.9726149975	8.0476350128	27.3503298477	2.49095797030667e-12	2.56786629763989e-09	18.7812837947
17	4.0863842046	6.8179929763	27.1776153087	2.68821261946175e-12	2.60819876149424e-09	18.2936307595
18	2.1224656023	9.1257597842	26.7231737323	3.29254738294199e-12	3.01707091856918e-09	18.5912448477
19	2.3036855426	8.7511497034	26.2447027108	4.09109163924757e-12	3.55149818409208e-09	18.3704623393
20	3.5993846125	7.5823264204	25.639686443	5.41376309389537e-12	4.46473042353551e-09	17.9529774646
21	4.2086199023	6.0797044863	25.3909754075	6.08580460089812e-12	4.77996481367683e-09	17.2193096815
22	2.2911077329	8.3518603011	25.2738438012	6.43306650404645e-12	4.82304540535191e-09	17.9090316524
23	2.7781059496	9.8310399895	24.7310107282	8.3470444662487e-12	5.9859196272307e-09	17.6178878145
24	3.2533287538	6.5940415535	24.3498376664	1.00554190328317e-11	6.91058673031361e-09	17.1523499222
25	-1.5423117594	10.1387338898	-23.6321132979	1.4388988004674e-11	9.4932787259637e-09	17.0597264833
26	2.2955363381	7.4764144583	21.6538192249	4.08971325788163e-11	2.49836038798146e-08	16.0531179152
27	2.77402164759	9.2450550735	-21.098178351	5.57493647321766e-11	3.28403579247329e-08	15.7485345494
28	3.044918854	7.461798952	20.8006577115	6.60199548830396e-11	3.75494184772709e-08	15.5812395495
29	ENSG00000180573	1.9632896038	7.6217999326	20.501827572	7.84231156794821e-11	4.31170290005793e-08
30	ENSG00000135821	2.3360642351	7.0970198514	20.3815296566	8.41076037725773e-11	4.45231385178948e-08
31	ENSG00000135549	-2.9795165369	7.8741107925	-20.3358901401	8.63793156646437e-11	4.45231385178948e-08
32	ENSG00000154122	2.2708346958	6.8408920316	20.1311552413	9.74173581161888e-11	4.86909668111642e-08
33	ENSG00000079691	2.6670365136	7.0155641091	19.7432538982	1.2273774701851e-10	5.66789168283092e-08
34	ENSG00000167034	2.5110559379	7.3105122891	19.7301634677	1.23708076016681e-10	5.66789168283092e-08
35	ENSG00000175928	-4.5186645689	5.7350091268	-19.6587591142	1.29146731140339e-10	5.75715184710472e-08
36	ENSG00000164597	1.8347233783	8.2929646766	19.5678765133	1.364453450633e-10	5.92244610914228e-08
37	ENSG00000135842	3.0135942622	6.3604321563	19.4663491508	1.45219059437999e-10	5.98809845820225e-08
38	ENSG00000077232	2.7675647372	6.9299618513	19.4653341341	1.45218830076446e-10	5.98809845820225e-08
39	ENSG000000213639	1.763750979	8.0133560104	19.3236548386	1.58357424925116e-10	6.27946786994643e-08
40	ENSG00000124225	2.9913518348	7.2036531829	19.3078669487	1.59899145469716e-10	6.27946786994643e-08
41	ENSG00000125257	2.8751935762	7.2202275479	19.0355008121	1.89212562317252e-10	7.25784186711804e-08
42	ENSG00000156802	2.5870454846	6.1344933012	18.8238113305	2.15998127143461e-10	8.09698433887329e-08
43	ENSG00000102172	2.2445597616	6.9300552353	18.7719170972	2.23171289925325e-10	8.1799716800629e-08
44	ENSG00000091879	4.5119520823	4.752223287	18.6691334125	2.3814923087883e-10	8.35138184381226e-08
45	ENSG00000143797	3.1908342549	5.818045836	18.6320449048	2.43817321394985e-10	8.35138184381226e-08
46	ENSG00000198648	2.4976444789	6.8059899323	18.6220042381	2.45376719951212e-10	8.35138184381226e-08

DAVID

Database for annotation, visualisation and integrated discovery

[Nat Protoc.](#) 2009;4(1):44-57. doi: 10.1038/nprot.2008.211.

Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.

Huang da W¹, Sherman BT, Lempicki RA.

Author information

Abstract

DAVID bioinformatics resources consists of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from large gene/protein lists. This protocol explains how to use DAVID, a high-throughput and integrated data-mining environment, to analyze gene lists derived from high-throughput genomic experiments. The procedure first requires uploading a gene list containing any number of common gene identifiers followed by analysis using one or more text and pathway-mining tools such as gene functional classification, functional annotation chart or clustering and functional annotation table. By following this protocol, investigators are able to gain an in-depth understanding of the biological themes in lists of genes that are enriched in genome-scale studies.

PMID: 19131956 [PubMed - indexed for MEDLINE]

- Very widely used (2012: ~1200 gene lists per day, 5000 different institutes)
- Easy to use, great for a quick look at the functional categories arising in a gene list
- Criticism about updating: But now DAVID 6.8 (current beta release) May. 2016

The screenshot shows the DAVID Functional Annotation tool's Analysis Wizard interface. On the left, a sidebar titled "Upload Gene List" contains fields for pasting a list or choosing from a file. Step 1: Enter Gene List (A) has a text area with IDs like ENSG00000138028, ENSG0000007516, etc. Step 2: Select Identifier (B) shows ENSEMBL_GENE_ID selected. Step 3: List Type (C) shows "Gene List" selected. Step 4: Submit List (D) contains a "Submit List" button. The main panel on the right is titled "Analysis Wizard" and provides instructions for Step 1. A red arrow points to the gene list input field with the text "Paste Ensembl Id here". Red arrows also point to the identifier selection and list type buttons with the text "Choose". A green arrow points to the submit button with the text "Choose".

Paste Ensembl Id here

Choose

Choose

Choose

Analysis Wizard

DAVID Bioinformatics Resources 6.7, NIAID/NIH

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

Upload List Background

Upload Gene List

Demolist 1 Demolist 2

Upload Help

Step 1: Enter Gene List

A: Paste a list

ENSG00000138028
ENSG0000007516
ENSG00000140691
ENSG00000070404
ENSG00000144354

Clear

Or

B: Choose From a File

Browse... 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

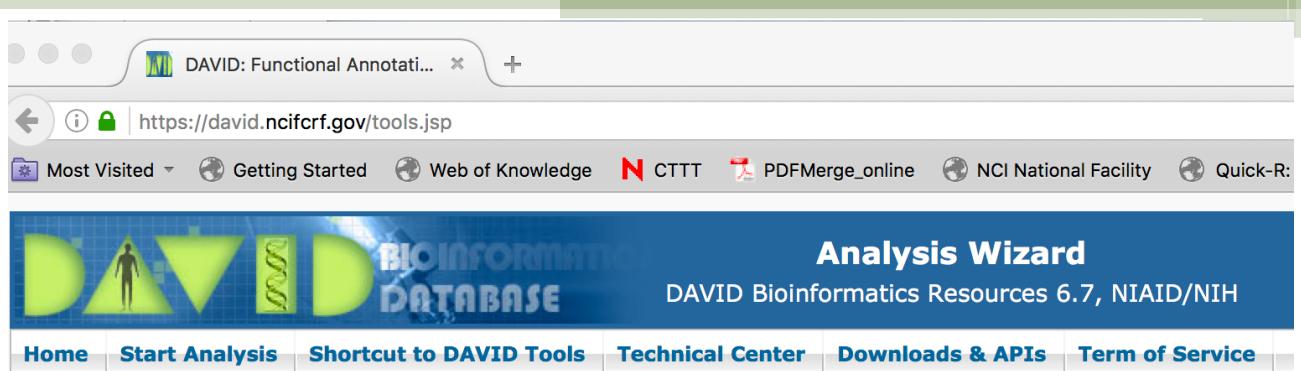
Tell us how you like the tool
Contact us for questions

Step 1. Submit your gene list through left panel.

An example:

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

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



The screenshot shows the DAVID Bioinformatics Resources 6.7, NIAID/NIH Analysis Wizard homepage. The URL is https://david.ncifcrf.gov/tools.jsp. The page features a blue header with the DAVID logo and navigation links for Home, Start Analysis, Shortcut to DAVID Tools, Technical Center, Downloads & APIs, and Term of Service.

Choose Homo sapiens



Gene List Manager

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

- Use All Species -
Homo sapiens(1402)
Unknown(1)

Select Species

List Manager Help

List_1

Select List to:
Use Rename
Remove Combine
Show Gene List

[View Unmapped Ids](#)

Choose



Choose



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



 [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

Functional Annotation Tool
DAVID Bioinformatics Resources 6.7, NIAID/NIH

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

Upload List Background

Gene List Manager

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

- Use All Species -
Homo sapiens(1402)
Unknown(1)

Select Species

List Manager Help

List_1

Select List to:
[Use](#) [Rename](#)
[Remove](#) [Combine](#)
[Show Gene List](#)
[View Unmapped Ids](#)

Annotation Summary Results

Current Gene List: List_1
Current Background: Homo sapiens
1392 DAVID IDs
Check Defaults Clear All

Uncheck Defaults

Expand

Select

Select

Annotation Summary Results

1392 DAVID IDs

Check Defaults Clear All

Disease (0 selected)

Functional_Categories (0 selected)

Gene_Ontology (1 selected)

GOTERM_BP_1 65.0% 905 Chart

GOTERM_BP_2 64.7% 900 Chart

GOTERM_BP_3 63.1% 878 Chart

GOTERM_BP_4 62.0% 863 Chart

GOTERM_BP_5 57.7% 803 Chart

GOTERM_BP_ALL 65.2% 908 Chart

GOTERM_BP_FAT 62.7% 873 Chart

GOTERM_CC_1 72.5% 1009 Chart

GOTERM_CC_2 70.5% 981 Chart

GOTERM_CC_3 70.5% 981 Chart

GOTERM_CC_4 67.7% 943 Chart

GOTERM_CC_5 66.5% 925 Chart

GOTERM_CC_ALL 72.5% 1009 Chart

GOTERM_CC_FAT 60.8% 847 Chart

GOTERM_MF_1 69.3% 964 Chart

GOTERM_MF_2 68.0% 946 Chart

GOTERM_MF_3 60.3% 839 Chart

GOTERM_MF_4 55.2% 768 Chart

GOTERM_MF_5 46.6% 648 Chart

GOTERM_MF_ALL 69.3% 964 Chart

GOTERM_MF_FAT 59.8% 832 Chart

PANTHER_BP_ALL 59.4% 827 Chart

PANTHER_MF_ALL 60.2% 838 Chart

General Annotations (0 selected)

Literature (0 selected)

Main_Accessions (0 selected)

Pathways (0 selected)

Protein_Domains (0 selected)

Protein_Interactions (0 selected)

Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering

Functional Annotation Chart

Functional Annotation Table

Functional Annotation Clustering

[Help and Manual](#)

Current Gene List: List_1

Current Background: Homo sapiens

1392 DAVID IDs

 Options Classification Stringency Medium
[Rerun using options](#)[Create Sublist](#)

157 Cluster(s)

[Download File](#)

Annotation Cluster 1		Enrichment Score: 8.99	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_FAT	cell_cycle_phase	RT		66	1.5E-11	4.9E-8
<input type="checkbox"/>	GOTERM_BP_FAT	cell_cycle_process	RT		80	3.7E-11	6.0E-8
<input type="checkbox"/>	GOTERM_BP_FAT	cell_cycle	RT		98	1.3E-10	1.4E-7
<input type="checkbox"/>	GOTERM_BP_FAT	nuclear_division	RT		43	1.6E-10	1.3E-7
<input type="checkbox"/>	GOTERM_BP_FAT	mitosis	RT		43	1.6E-10	1.3E-7
<input type="checkbox"/>	GOTERM_BP_FAT	M_phase	RT		55	1.6E-10	1.0E-7
<input type="checkbox"/>	GOTERM_BP_FAT	M_phase_of_mitotic_cell_cycle	RT		43	2.8E-10	1.5E-7
<input type="checkbox"/>	GOTERM_BP_FAT	organelle_fission	RT		43	5.7E-10	2.6E-7
<input type="checkbox"/>	GOTERM_BP_FAT	mitotic_cell_cycle	RT		57	1.7E-9	6.7E-7
<input type="checkbox"/>	GOTERM_BP_FAT	chromosome_segregation	RT		19	3.0E-6	9.6E-4
<input type="checkbox"/>	GOTERM_BP_FAT	cell_division	RT		41	6.2E-6	1.8E-3
Annotation Cluster 2		Enrichment Score: 2.74	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_FAT	chromosome_segregation	RT		19	3.0E-6	9.6E-4
<input type="checkbox"/>	GOTERM_BP_FAT	mitotic_sister_chromatid_segregation	RT		11	6.4E-5	1.7E-2
<input type="checkbox"/>	GOTERM_BP_FAT	sister_chromatid_segregation	RT		11	8.3E-5	2.0E-2
<input type="checkbox"/>	GOTERM_BP_FAT	mitotic_chromosome_condensation	RT		4	4.7E-2	6.9E-1
<input type="checkbox"/>	GOTERM_BP_FAT	chromosome_condensation	RT		5	7.4E-2	7.6E-1
<input type="checkbox"/>	GOTERM_BP_FAT	chromosome_organization	RT		30	6.9E-1	1.0E0
Annotation Cluster 3		Enrichment Score: 2.55	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_FAT	response_to_organic_substance	RT		71	3.6E-4	6.6E-2
<input type="checkbox"/>	GOTERM_BP_FAT	response_to_endogenous_stimulus	RT		40	7.8E-3	3.4E-1
<input type="checkbox"/>	GOTERM_BP_FAT	response_to_hormone_stimulus	RT		37	8.0E-3	3.3E-1
Annotation Cluster 4		Enrichment Score: 2.51	G		Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_FAT	regulation_of_cell_cycle	RT		46	1.5E-6	5.5E-4
<input type="checkbox"/>	GOTERM_BP_FAT	regulation_of_mitotic_cell_cycle	RT		24	1.1E-4	2.5E-2
<input type="checkbox"/>	GOTERM_BP_FAT	regulation_of_cell_cycle_process	RT		19	3.5E-4	6.8E-2
<input type="checkbox"/>	GOTERM_BP_FAT	regulation_of_mitosis	RT		12	7.7E-4	1.1E-1
<input type="checkbox"/>	GOTERM_BP_FAT	regulation_of_nuclear_division	RT		12	7.7E-4	1.1E-1
<input type="checkbox"/>	GOTERM_BP_FAT	positive_regulation_of_cell_cycle	RT		11	3.2E-3	2.6E-1

What DEGs have this annotation?

Functional Annotation Clustering

[Help and Manual](#)

Current Gene List: List_1

Current Background: Homo sapiens

1392 DAVID IDs

 Options Classification Stringency Medium[Rerun using options](#)[Create Sublist](#)

157 Cluster(s)

[Download File](#)

Annotation Cluster 1	Enrichment Score: 8.99	G		Count	P_Value	Benjamini
<input type="checkbox"/> GOTERM_BP_FAT	cell_cycle_phase	RT		66	1.5E-11	4.9E-8
<input type="checkbox"/> GOTERM_BP_FAT	cell_cycle_process	RT		80	3.7E-11	6.0E-8
<input type="checkbox"/> GOTERM_BP_FAT	cell_cycle	RT		98	1.3E-10	1.4E-7
<input type="checkbox"/> GOTERM_BP_FAT	nuclear_division	RT		43	1.6E-10	1.3E-7
<input type="checkbox"/> GOTERM_BP_FAT	mitosis	RT		43	1.6E-10	1.3E-7
<input type="checkbox"/> GOTERM_BP_FAT	M_phase	RT		55	1.6E-10	1.0E-7
<input type="checkbox"/> GOTERM_BP_FAT	M_phase_of_mitotic_cell_cycle	RT		43	2.8E-10	1.5E-7
<input type="checkbox"/> GOTERM_BP_FAT	organelle_fission	RT		43	5.7E-10	2.6E-7
<input type="checkbox"/> GOTERM_BP_FAT	mitotic_cell_cycle	RT		57	1.7E-9	6.7E-7
<input type="checkbox"/> GOTERM_BP_FAT	chromosome_segregation	RT		19	3.0E-6	9.6E-4
<input type="checkbox"/> GOTERM_BP_FAT	cell_division	RT		41	6.2E-6	1.8E-3
Annotation Cluster 2	Enrichment Score: 2.74	G		Count	P_Value	Benjamini
<input type="checkbox"/> GOTERM_BP_FAT	chromosome_segregation	RT		19	3.0E-6	9.6E-4
<input type="checkbox"/> GOTERM_BP_FAT	mitotic_sister_chromatid_segregation	RT		11	6.4E-5	1.7E-2
<input type="checkbox"/> GOTERM_BP_FAT	sister_chromatid_segregation	RT		11	8.3E-5	2.0E-2
<input type="checkbox"/> GOTERM_BP_FAT	mitotic_chromosome_condensation	RT		4	4.7E-2	6.9E-1
<input type="checkbox"/> GOTERM_BP_FAT	chromosome_condensation	RT		5	7.4E-2	7.6E-1
<input type="checkbox"/> GOTERM_BP_FAT	chromosome_organization	RT		30	6.9E-1	1.0E0
Annotation Cluster 3	Enrichment Score: 2.55	G		Count	P_Value	Benjamini
<input type="checkbox"/> GOTERM_BP_FAT	response_to_organic_substance	RT		71	3.6E-4	6.6E-2
<input type="checkbox"/> GOTERM_BP_FAT	response_to_endogenous_stimulus	RT		40	7.8E-3	3.4E-1
<input type="checkbox"/> GOTERM_BP_FAT	response_to_hormone_stimulus	RT		37	8.0E-3	3.3E-1
Annotation Cluster 4	Enrichment Score: 2.51	G		Count	P_Value	Benjamini
<input type="checkbox"/> GOTERM_BP_FAT	regulation_of_cell_cycle	RT		46	1.5E-6	5.5E-4
<input type="checkbox"/> GOTERM_BP_FAT	regulation_of_mitotic_cell_cycle	RT		24	1.1E-4	2.5E-2
<input type="checkbox"/> GOTERM_BP_FAT	regulation_of_cell_cycle_process	RT		19	3.5E-4	6.8E-2
<input type="checkbox"/> GOTERM_BP_FAT	regulation_of_mitosis	RT		12	7.7E-4	1.1E-1
<input type="checkbox"/> GOTERM_BP_FAT	regulation_of_nuclear_division	RT		12	7.7E-4	1.1E-1
<input type="checkbox"/> GOTERM_BP_FAT	positive_regulation_of_cell_cycle	RT		11	3.2E-3	2.6E-1

Look at genes which have these ontology terms in common

Gene ontology terms



Genes

DAVID

- Spend some time looking at the clusters, and the genes in the clusters
- Does it seem sensible that genes in these functional categories would be dysregulated in this experiment?
- Go back to the Annotation Summary and select all the defaults and run again – what clusters are coming up now?

GORilla

Gene ontology enrichment analysis and visualization

BMC Bioinformatics



Software

Open Access

GORilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists

Eran Eden*^{†1}, Roy Navon*^{†2,3}, Israel Steinfield^{3,4}, Doron Lipson⁴ and Zohar Yakhini*^{3,4}

Address: ¹Molecular Cell Biology Department, Weizmann Institute of Science, Rehovot, Israel, ²School of Computer Science, Tel Aviv University, Tel Aviv, Israel, ³Agilent Laboratories, Tel-Aviv, Israel and ⁴Computer Science Department, Technion, Haifa, Israel

Email: Eran Eden* - eraneden@gmail.com; Roy Navon* - rnavon@gmail.com; Israel Steinfield - israels@technion.ac.il; Doron Lipson - dlipson@cs.technion.ac.il; Zohar Yakhini* - zohar_yakhini@agilent.com

* Corresponding authors †Equal contributors

Published: 3 February 2009

BMC Bioinformatics 2009, **10**:48 doi:10.1186/1471-2105-10-48

This article is available from: <http://www.biomedcentral.com/1471-2105/10/48>

Received: 16 October 2008

Accepted: 3 February 2009

- Allows use of a single ranked list of genes
- Or list of DEGs and background
- Shows hierachal structure of the ontology
- Can use output directly with REVIGO



Gene Ontology enRICHment anaLysis and visualLizAtion tool

<http://cbl-gorilla.cs.technion.ac.il>

GORILLA is a tool for identifying and visualizing enriched GO terms in ranked lists of genes.
It can be run in one of two modes:

1. Searching for enriched GO terms that appear densely at the top of a ranked list of genes or
2. Searching for enriched GO terms in a target list of genes compared to a background list of genes.

For further details see [References](#).

[Running example](#)

[Usage instructions](#)

[GORILLA News](#)(Updated March 8th 2013)

[References](#)

Step 1: Choose organism

Step 2: Choose running mode

Single ranked list of genes Two unranked lists of genes (target and background lists)

Step 3: Paste a ranked list of gene/protein names

Names should be separated by an <ENTER>. The preferred format is gene symbol.
Other supported formats are: gene and protein RefSeq, Uniprot, Unigene and
Ensembl. Use [WebGestalt](#) for conversion from other identifier formats.

Target set:

```
ENSG000000116041  
ENSG000000121966  
ENSG000000187559  
ENSG000000196872  
ENSG000000065989  
ENSG000000179967  
ENSG000000221164
```

Or upload a file: no file selected

Background set:

```
ENSG000000201821  
ENSG000000217264  
ENSG000000209424  
ENSG000000219008  
ENSG000000212553  
ENSG000000212559
```

Or upload a file: no file selected

Step 4: Choose an ontology

Process Function Component All

Select Process

Advanced parameters

P-value threshold:

Analysis name: (optional)

E-mail address: (optional - enter an e-mail address if you would like to receive a link to your results)

Output results in Microsoft Excel format

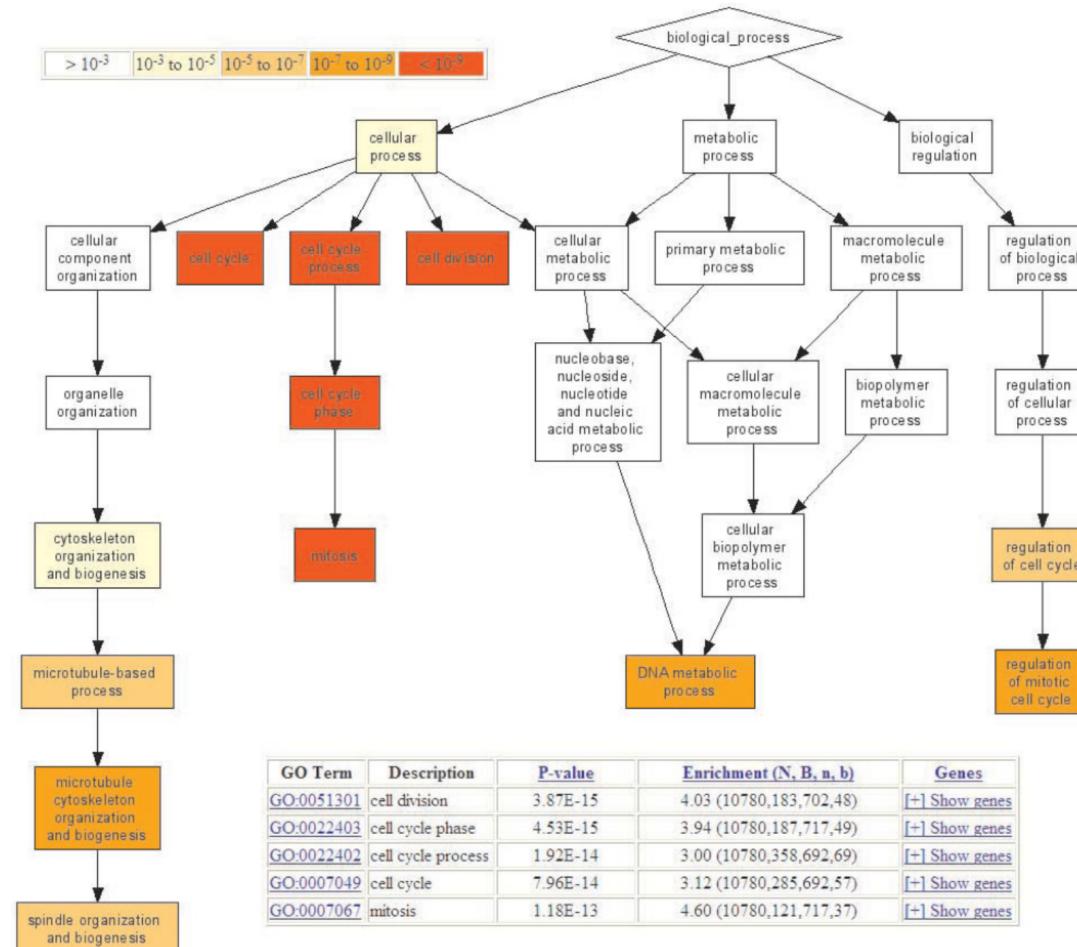
Include unresolved and duplicate genes in output

Show output also in [REVIGO](#)

Paste Ensembl Ids here
voom_res_sig_lfc.txt

Paste Ensembl Ids here
voom_res.txt

Choose REVIGO

**Figure 2**

An example of the GOrilla analysis output. 14,565 genes from the van't Veer dataset were ranked according to their differential expression and given as input to GOrilla. The resulting enriched GO terms are visualized using a DAG graphical representation with color coding reflecting their degree of enrichment. Nodes in the graph are clickable and give additional information on the GO terms and genes attributing to the enrichment. N is the total number of genes; B is the total number of genes associated with a specific GO term; n is the flexible cutoff, i.e. the automatically determined number of genes in the 'target set' and b is the number of genes in the 'target set' that are associated with a specific GO term. Enrichment is defined as $(b/n)/(B/N)$.

Gorilla enriched GO terms

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0044763	single-organism cellular process	7.68E-14	1.07E-9	1.21 (16315,8612,1051,670)	[+] Show genes
GO:0044699	single-organism process	1.27E-10	8.89E-7	1.15 (16315,10021,1051,741)	[+] Show genes
GO:0051726	regulation of cell cycle	2.9E-10	1.35E-6	1.90 (16315,818,1051,100)	[+] Show genes
GO:0022402	cell cycle process	3.48E-9	1.21E-5	1.75 (16315,985,1051,111)	[+] Show genes
GO:1903047	mitotic cell cycle process	3.65E-9	1.02E-5	1.93 (16315,675,1051,84)	[+] Show genes
GO:0019220	regulation of phosphate metabolic process	2.54E-8	5.92E-5	1.56 (16315,1454,1051,146)	[+] Show genes
GO:0042325	regulation of phosphorylation	4.13E-8	8.23E-5	1.60 (16315,1239,1051,128)	[+] Show genes
GO:0051174	regulation of phosphorus metabolic process	4.48E-8	7.82E-5	1.54 (16315,1467,1051,146)	[+] Show genes
GO:0001932	regulation of protein phosphorylation	1.44E-7	2.24E-4	1.60 (16315,1154,1051,119)	[+] Show genes
GO:0007346	regulation of mitotic cell cycle	2.18E-7	3.04E-4	2.12 (16315,381,1051,52)	[+] Show genes
GO:0010564	regulation of cell cycle process	4E-7	5.07E-4	1.97 (16315,464,1051,59)	[+] Show genes
GO:0045786	negative regulation of cell cycle	5.8E-7	6.74E-4	2.05 (16315,393,1051,52)	[+] Show genes
GO:0010033	response to organic substance	6.19E-7	6.65E-4	1.46 (16315,1636,1051,154)	[+] Show genes
GO:0031399	regulation of protein modification process	1.54E-6	1.53E-3	1.46 (16315,1530,1051,144)	[+] Show genes
GO:1901990	regulation of mitotic cell cycle phase transition	1.59E-6	1.48E-3	2.37 (16315,229,1051,35)	[+] Show genes
GO:0043067	regulation of programmed cell death	1.77E-6	1.55E-3	1.52 (16315,1224,1051,120)	[+] Show genes
GO:0042127	regulation of cell proliferation	1.89E-6	1.55E-3	1.50 (16315,1328,1051,128)	[+] Show genes
GO:0048523	negative regulation of cellular process	2.03E-6	1.57E-3	1.26 (16315,3745,1051,304)	[+] Show genes
GO:0042981	regulation of apoptotic process	2.15E-6	1.58E-3	1.52 (16315,1216,1051,119)	[+] Show genes
GO:0051302	regulation of cell division	2.64E-6	1.84E-3	2.29 (16315,244,1051,36)	[+] Show genes
GO:0048519	negative regulation of biological process	3.06E-6	2.03E-3	1.24 (16315,4007,1051,321)	[+] Show genes

Have a look at some of these genes

GO:00000703	protein phosphorylation	7.34E-4	7.12E-2	1.44 (16315,56,1051,11)	[+] Show genes
GO:2000648	positive regulation of stem cell proliferation	7.67E-4	7.39E-2	3.05 (16315,56,1051,11)	[+] Show genes
GO:0033045	regulation of sister chromatid segregation	7.67E-4	7.34E-2	3.05 (16315,56,1051,11)	[+] Show genes
GO:0000281	mitotic cytokinesis	7.76E-4	7.37E-2	4.35 (16315,25,1051,7)	[+] Show genes
GO:0002053	positive regulation of mesenchymal cell proliferation	7.76E-4	7.32E-2	4.35 (16315,25,1051,7)	[+] Show genes
GO:0045667	regulation of osteoblast differentiation	7.86E-4	7.36E-2	2.44 (16315,102,1051,16)	[+] Show genes
GO:0051270	regulation of cellular component movement	7.93E-4	7.38E-2	1.48 (16315,690,1051,66)	[+] Show genes
GO:0009987	cellular process	8.57E-4	7.92E-2	1.06 (16315,11669,1051,796)	[+] Show genes
GO:0010959	regulation of metal ion transport	8.57E-4	7.87E-2	1.81 (16315,275,1051,32)	[+] Show genes
GO:0007062	sister chromatid cohesion	8.76E-4	7.99E-2	2.41 (16315,103,1051,16)	[+] Show genes
GO:0044702	single organism reproductive process	8.83E-4	8E-2	1.38 (16315,1020,1051,91)	[+] Show genes
GO:0006915	apoptotic process	9.04E-4	8.15E-2	1.57 (16315,504,1051,51)	[+] Show genes
GO:0032331	negative regulation of chondrocyte differentiation	9.23E-4	8.26E-2	4.90 (16315,19,1051,6)	[+] Show genes
GO:0044710	single-organism metabolic process	9.34E-4	8.31E-2	1.17 (16315,3783,1051,286)	[+] Show genes
GO:0006695	cholesterol biosynthetic process	9.42E-4	8.33E-2	3.76 (16315,33,1051,8)	[+] Show genes
GO:1902653	secondary alcohol biosynthetic process	9.42E-4	8.27E-2	3.76 (16315,33,1051,8)	[+] Show genes
GO:0010948	negative regulation of cell cycle process	9.63E-4	8.4E-2	1.96 (16315,198,1051,25)	[+] Show genes
GO:0006576	cellular biogenic amine metabolic process	9.84E-4	8.53E-2	2.31 (16315,114,1051,17)	[+] Show genes
GO:0044106	cellular amino metabolic process	9.84E-4	8.48E-2	2.31 (16315,114,1051,17)	[+] Show genes
GO:0030071	regulation of mitotic metaphase/anaphase transition	9.92E-4	8.5E-2	3.41 (16315,41,1051,9)	[+] Show genes

Note that only 42.01% of the terms you entered were used in the analysis

Species used: Homo sapiens

The system has recognized 19562 genes out of 38837 gene terms entered by the user.

0 genes were recognized by gene symbol and 19562 genes by other gene IDs .

1118 duplicate genes were removed (keeping the highest ranking instance of each gene) leaving a total of 18444 genes.
Only 16315 of these genes are associated with a GO term.

[Output in Microsoft Excel format](#)

[Visualize output in REViGO](#)

The GOrilla database is periodically updated using the [GO database](#) and other sources.

The GOrilla database was last updated on May 7, 2016

This results page will be available on this site for one month from now (until Jun 8, 2016). You can bookmark this page and come back to it later.

'P-value' is the enrichment p-value computed according to the mHG or HG model. This p-value is not corrected for multiple testing of 13961 GO terms.

'FDR q-value' is the correction of the above p-value for multiple testing using the Benjamini and Hochberg (1995) method.
Namely, for the i^{th} term (ranked according to p-value) the FDR q-value is $(\text{p-value} * \text{number of GO terms}) / i$.

Enrichment (N, B, n, b) is defined as follows:

N - is the total number of genes

B - is the total number of genes associated with a specific GO term

n - is the number of genes in the top of the user's input list or in the target set when appropriate

b - is the number of genes in the intersection

Enrichment = $(b/n) / (B/N)$

Genes: For each GO term you can see the list of associated genes that appear in the optimal top of the list.
Each gene name is specified by gene symbol followed by a short description of the gene

Select REViGO



REVIGO

Reduce and visualize gene ontology

OPEN  ACCESS Freely available online



REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms

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

1 Division of Electronics, Rudjer Boskovic Institute, Zagreb, Croatia, **2** Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG) and UPF, Barcelona, Spain

Abstract

Outcomes of high-throughput biological experiments are typically interpreted by statistical testing for enriched gene functional categories defined by the Gene Ontology (GO). The resulting lists of GO terms may be large and highly redundant, and thus difficult to interpret. REVIGO is a Web server that summarizes long, unintelligible lists of GO terms by finding a representative subset of the terms using a simple clustering algorithm that relies on semantic similarity measures. Furthermore, REVIGO visualizes this non-redundant GO term set in multiple ways to assist in interpretation: multidimensional scaling and graph-based visualizations accurately render the subdivisions and the semantic relationships in the data, while treemaps and tag clouds are also offered as alternative views. REVIGO is freely available at <http://revigo.irb.hr/>.

Citation: Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE 6(7): e21800. doi:10.1371/journal.pone.0021800

- Summarizes lists of GO terms reducing redundancy
- Visualization tools – can download R script for making treemap

<http://revigo.irb.hr>

Results REVIGO summarizes and visualizes long lists of Gene Ontology terms

REVIGO

reduce + visualize Gene ontology

Welcome to REVIGO!

REVIGO can take long lists of Gene Ontology terms and summarize them by removing redundant GO terms. The remaining terms can be visualized in semantic similarity-based scatterplots, interactive graphs, or tag clouds. [More about REVIGO...](#) | [?](#)

Please enter a list of Gene Ontology IDs below, each on its own line. The GO IDs may be followed by p-values or another quantity which describes the GO term in a way meaningful to you. [?](#)

Examples: #1 #2 #3

```
% List generated using Gorilla  
% http://cbl-gorilla.cs.technion.ac.il/  
  
% GO term pValue  
GO:0044763 7.68E-14  
GO:0044699 1.27E-10  
GO:0051726 2.9E-10  
GO:0022402 3.48E-9  
GO:1903047 3.65E-9  
GO:0019220 2.54E-8  
GO:0042325 4.13E-8  
GO:0051174 4.48E-8  
GO:0001932 1.44E-7  
GO:0007346 2.18E-7  
GO:0010564 4E-7  
GO:0045786 5.8E-7  
GO:0010033 6.19E-7  
GO:0031399 1.54E-6  
GO:1901990 1.59E-6  
GO:0043067 1.77E-6
```

Allowed similarity: How large would you like the resulting list to be?
 Large (allowed similarity=0.9) Medium (0.7) Small (0.5) Tiny (0.4) [?](#)

If provided, the numbers associated to GO categories are...
 p-values
 some other quantity, where [?](#)

Advanced options:

Select a database with GO term sizes: [?](#)

Select a semantic similarity measure to use: [?](#)

The version of the Gene Ontology used is monthly release of Oct 2014 "[go_201410-termdb.obo-xml.gz](#)". The UniProt-to-GO mapping file "[gene_association.goa_uniprot.gz](#)" is dated 30th Sep 2014, downloaded from the EBI GOA project.

If you found REVIGO useful in your work, please cite the following reference:
Supek F, Bošnjak M, Škunca N, Šmuc T.

Q+A

Q: I have a list of interesting genes, but not a list of GO terms.

A: You can use one of the following web servers to search for GO terms that are overrepresented in your list of genes:

- [GOrilla](#) (multiple eukaryotes)
- [AgriGO](#) (many plants, several animals)
- [L2L](#) (human, mouse and rat)
- [GOTermFinder](#) (many species)
- [FatGO](#) (multiple eukaryotes)

After that, return here with the list of GO terms and p-values (or enrichments).

Q: I still don't have a list of interesting genes, but I'd like to try out my favourite GO enrichment tool and then bring the output to REVIGO to summarize and visualize.

A: Here are the links to two example gene sets, one from [agriGO](#) (click "Example") and another one from [DAVID](#) (click "Demolist_1").

Q: The organism I work on is not listed in the "GO term sizes" box in Advanced options.

A: The chosen database is used to find the size of each GO term i.e. the percentage of genes annotated with the term. This quantity determines the size of bubbles in the visualizations, thus indicating a more general GO term (larger) or a more specific one (smaller). The choice of database also has some influence on the GO term clustering/selection process, and on the bubble placement in the visualizations. If your organism is not available, select the closest relative, e.g. human or mouse should work for any mammal. The default choice (whole UniProt) should also suffice in most cases.

Q: I have made a Web server/software which produces lists of GO categories. Can its output be fed into REVIGO?

A: Yes! REVIGO can be used to summarize and visualize the results of your server. Please see instructions on [this page](#), or email [Fran Supek](#) for more information.

REVIGO is an iProject funded by the Ministry of Science, Education and Sport of Croatia (2008-057) implemented at the Laboratory for information systems, at the Rudjer Boskovic Institute, Croatia.

Other iProjects at the Rudjer Boskovic Institute:
GORBI - function predictions for thousands of poorly characterized bacterial proteins using phylogenetic profiling.

REVIGO

Biological Process (162)

Tag Clouds

Scatterplot & Table

Interactive Graph

TreeMap

Choose Treemap



While parsing your data, warning(s) were encountered:

Go term 98813 was not found in the current version of the GeneOntology, dated 29:09:2014 15:10. GO term will be skipped.

Tip: your resulting list of GO terms seems to be quite long. If you want to reduce it further, press the Back button in your browser and choose a different setting for the "allowed similarity" parameter.

REVIGO

Biological Process (94)

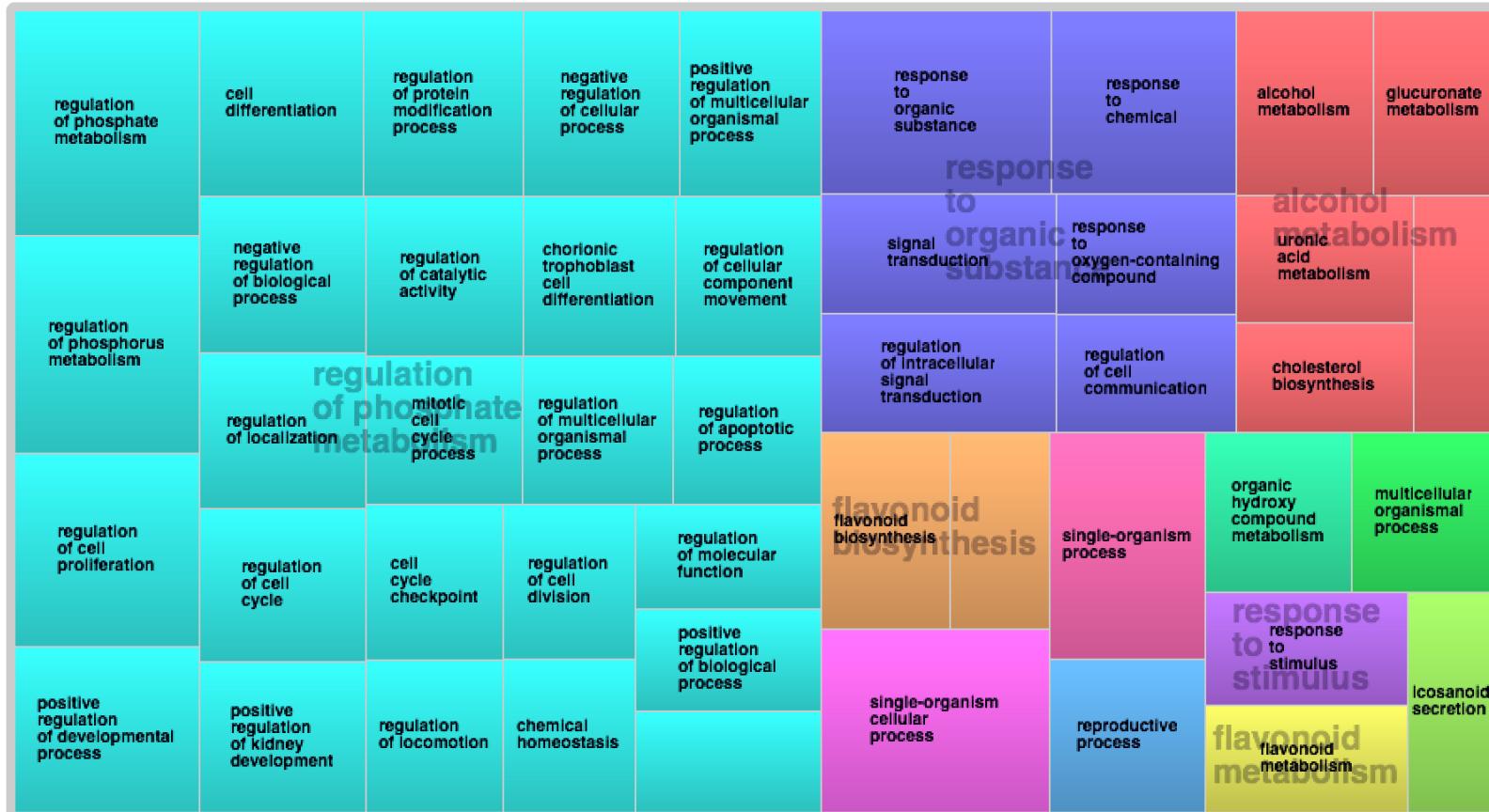
Tag Clouds

Scatterplot & Table

Interactive Graph

TreeMap

Show
 abs_log10_pvalue
 size



[Export data to text table \(CSV\)](#)



[Make R script for plotting treemaps](#)

by DrasticData

Published online 28 October 2014

Nucleic Acids Research, 2015, Vol. 43, Database issue **D447–D452**
doi: 10.1093/nar/gku1003

STRING v10: protein–protein interaction networks, integrated over the tree of life

Damian Szklarczyk¹, Andrea Franceschini¹, Stefan Wyder¹, Kristoffer Forslund²,
Davide Heller¹, Jaime Huerta-Cepas², Milan Simonovic¹, Alexander Roth¹, Alberto Santos³,
Kalliopi P. Tsafou³, Michael Kuhn^{4,5}, Peer Bork^{2,*}, Lars J. Jensen^{3,*} and
Christian von Mering^{1,*}

¹Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, 8057 Zurich, Switzerland, ²European Molecular Biology Laboratory, 69117 Heidelberg, Germany, ³Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, 2200 Copenhagen N, Denmark, ⁴Biotechnology Center, Technische Universität Dresden, 01062 Dresden, Germany and ⁵Max Planck Institute of Molecular Cell Biology and Genetics, 01062 Dresden, Germany

- Web tool and also a Bioconductor package (String.db)
- Builds protein-protein interaction network from gene list
- Also allows GO and KEGG enrichment analysis of the network

Version: 10.0

LOGIN



Search

Download

Help

M

<http://string-db.org>

Paste top 500
Ensembl Ids

Protein by name >

Protein by sequence >

Multiple proteins >

Multiple sequences >

Organisms >

Protein families ("COGs") >

Examples >

Random entry >

SEARCH

Multiple Proteins by Names / Identifiers

List Of Names: (one per line; examples: #1 #2 #3)

```
ENSG00000214814
ENSG00000197971
ENSG00000068745
ENSG00000100528
ENSG00000003989
ENSG00000054690
```

... or, upload a file:

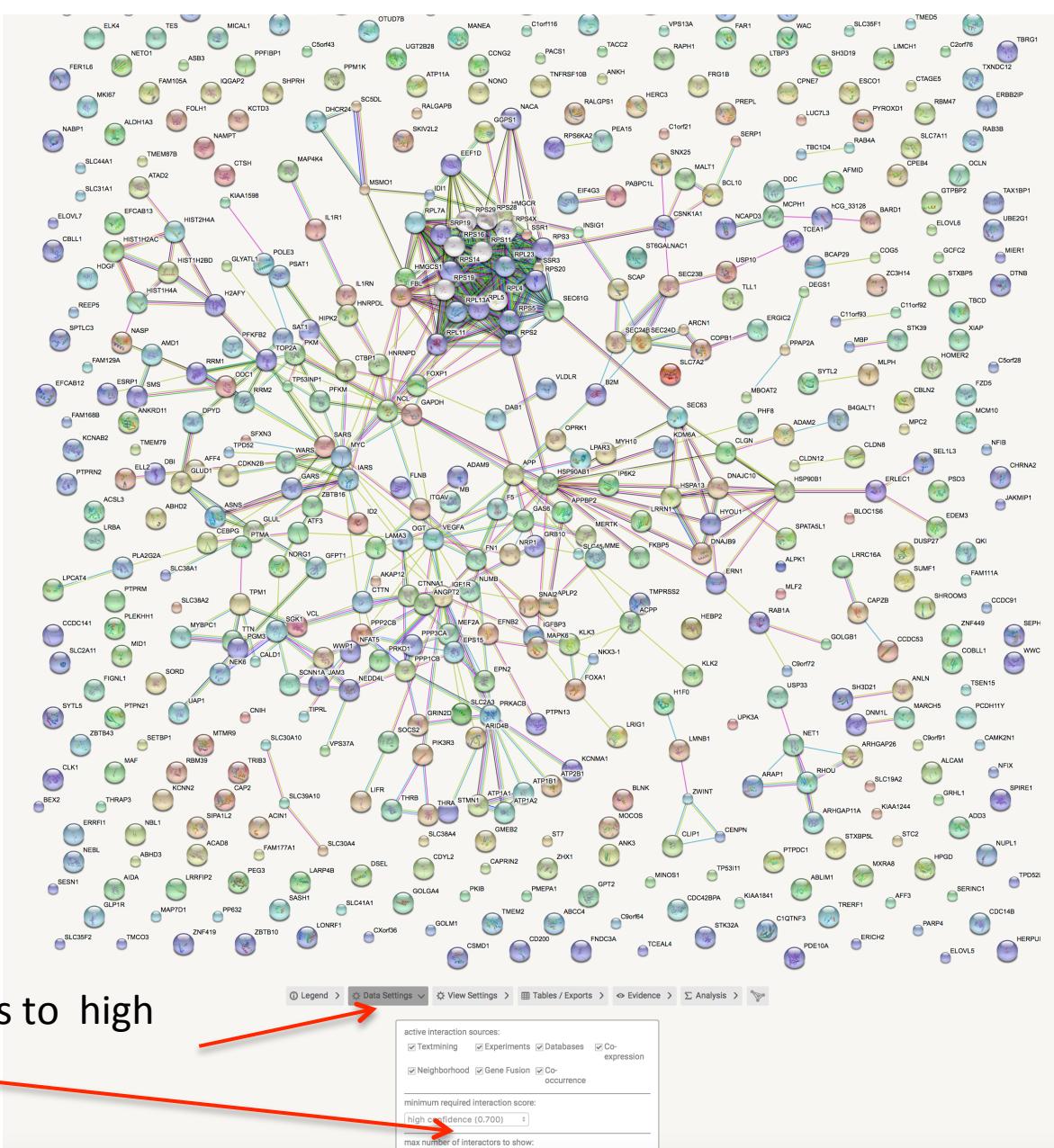
Organism:

Homo sapiens ▾

Select Homo sapiens

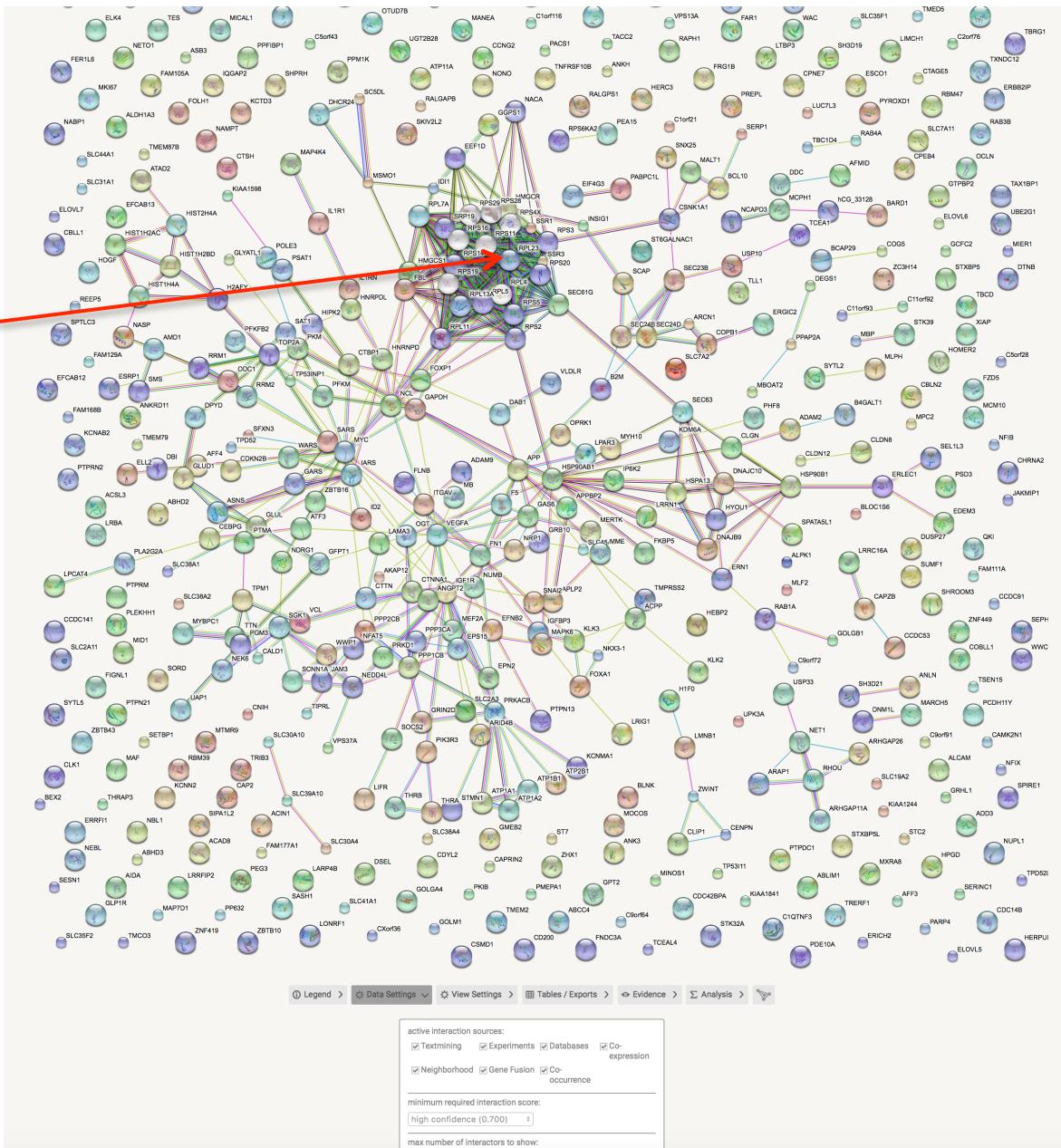
Select

SEARCH



Change default settings to high confidence

Click on highly connected proteins, such as in this cluster around CDK1 to find out more about their function



Summary

- Many tools are available for GO, and pathway analysis. We have presented a few here.
- Look at whether tools use up-to-date databases
- Use more than one tool
- Be aware that databases are collected from a broad environment - consider the relevance for your particular experiment