

Gene Set Analysis –Methods and Tools

Exercise 2.1

Antonio Mora

Lv Xuanyi

(Xie Chengshu)

December, 2019

Exercise 1. Using DAVID

DAVID is the acronym for “The Database for Annotation, Visualization and Integrated Discovery”. You can find it at: <https://david.ncifcrf.gov/home.jsp>

The picture below is its main page, which contains some general information about this platform. DAVID provides four main tools (details on the website):

1. Functional Annotation
2. Gene Functional Classification
3. Gene ID Conversion
4. Gene Name Batch Viewer

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

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

Welcome to DAVID 6.8

2003 - 2018

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

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

What's Important in DAVID?

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

Statistics of DAVID

DAVID Citations (2003-2017)

Year	Citations
03	0
04	0
05	0
06	0
07	0
08	0
09	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	4629

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

Screen Shot 1 Screen Shot 2 Screen Shot 3

1. Upload datasets

Click on the “Start Analysis” button.

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

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

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

Welcome to DAVID 6.8

2003 - 2018

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

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

What's Important in DAVID?

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

Statistics of DAVID

DAVID Citations (2003-2017)

Year	Citations
03	~100
04	~200
05	~300
06	~400
07	~500
08	~600
09	~700
10	~800
11	~900
12	~1000
13	~1100
14	~1200
15	~1300
16	~1400
17	~1500

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

Screen Shot 1 Screen Shot 2 Screen Shot 3

On the left panel of the page, there will be 3 steps:

1. Paste the gene list or choose a gene list file to upload. There are two ways to upload your gene list. One is to load a gene list from a file, another is to paste a gene list to the text box. Here we can upload the "affy_id.txt" file. Regarding the limitations of gene lists, please see DAVID FAQs. (<https://david.ncifcrf.gov/content.jsp?file=FAQs.html>).

2. Select the ID format, according to the format of the gene list. Here we use "affymetrix ID".

3. The list type may be a gene list or using a list as background. We choose the "gene list".

At last, click on the "Submit List" button.

The screenshot shows the DAVID Analysis Wizard interface. The browser address bar displays <https://david.ncifcrf.gov/tools.jsp>. The page header includes the DAVID Bioinformatics Resources 6.8, NIAID/NIH logo and navigation links: Home, Start Analysis, Shortcut to DAVID Tools, Technical Center, Downloads & APIs, Term of Service, Why DAVID?, and About Us.

A red box highlights the left panel, which contains the following steps:

- Upload Gene List**
 - Demolist 1 Demolist 2
 - Upload Help
 - Step 1: Enter Gene List
 - A: Paste a list (text area with a list of IDs and a Clear button)
 - Or
 - B: Choose From a File (Browse... affy_id.txt)
 - Multi-List File ?
 - Step 2: Select Identifier (Dropdown menu: AFFYMETRIX_3PRIME_IVT_ID)
 - Step 3: List Type (Radio buttons: Gene List (selected), Background)
 - Step 4: Submit List (Submit List button)

The main content area displays the title "Analysis Wizard" and a link: [Tell us how you like the tool](#) / [Contact us for questions](#). Below this, an arrow points to the text: "Step 1. Submit your gene list through left panel." An example gene list is provided:

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
```

2. Use DAVID tools

After task submission, the left panel shows the summary of the submitted gene list. The different available tools can be found under “Step 2”.

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

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](#)

Analysis tool types

Tool 1. ID conversion

We can click on the “Gene ID Conversion Tool”, go to the new page, and select a new ID format (Entrez_Gene_ID). In the left panel we find that there are 2499 genes from our uploaded gene list that can be found in the DAVID database, and 1 that cannot be found. Click on the “Submit to conversion tool” button.

https://david.ncicrf.gov/conversion.jsp

Gene ID Conversion Tool

DAVID Bioinformatics Resources 6.8, NIAID/NIH

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

*** Welcome to DAVID 6.8 ***
*** If you are looking for DAVID 6.7, please visit our development site. ***

Gene ID Conversion Tool

[Help and Tool Manual](#)

Upload List Background

Gene List Manager

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

- Use All Species -
Homo sapiens(2499)
Unknown(1)

Select Species

List Manager [Help](#)

List_1

Select List to:

Use Rename
Remove Combine
Show Gene List

[View Unmapped Ids](#)

Option 1: Convert the gene list being selected in left panel to

Option 2:

On the left side of the ID conversion result page, there is a summary table of the gene list conversion. In the table, there are 2499 affymetrix IDs converted into Entrez Gene IDs in DAVID database. On the right top corner, a “download file” option allows to download the whole conversion file.

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

[Save results](#)

Gene Accession Conversion Tool

[Help](#)
[Download File](#)

Gene Accession Conversion Statistics			Submit Converted List to DAVID as a Gene List				Submit Converted List to DAVID as a Background						
Conversion Summary							David Gene Name						
ID Count	In DAVID DB	Conversion	From	To	Species								
2499	Yes	Successful	201903_at	7384	Homo sapiens	ubiquinol-cytochrome c reductase core protein I(UQCRC1)							
0	Yes	None	203765_at	25801	Homo sapiens	grancalcin(GCA)							
0	No	None	201193_at	341	Homo sapiens	isocitrate dehydrogenase (NADP(+)) 1, cytosolic(IDH1)							
0	Ambiguous	Pending	203254_s_at	7094	Homo sapiens	talin 1(TLN1)							
Total Unique User IDs: 2499													
Summary of Ambiguous Gene IDs													
ID Count	Possible Source	Convert All											
All Possible Sources For Ambiguous IDs													
Ambiguous ID	Possibility	Convert											
202614_at	10463						solute carrier family 30 member 9(SLC30A9)						
202602_s_at	27336						HIV-1 Tat specific factor 1(HTATSF1)						
201196_s_at	262						adenosylmethionine decarboxylase 1(AMD1)						
201746_at	7157						tumor protein p53(TP53)						
201141_at	10457						glycoprotein nmb(GPNMB)						
202215_s_at	4802						nuclear transcription factor Y subunit gamma(NFYC)						

Tool 2. Gene Name Batch Viewer

This tool converts gene list IDs into gene names directly. Click the “Gene Name Batch Viewer” under the list of “Shortcut to DAVID tools”.

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

Upload List Background

Gene List Manager

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

- Use All Species -
Homo sapiens(2499)
Unknown(1)

Select Species

List Manager [Help](#)

List_1

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

Gene List Report

Current Gene List: List_1
Current Background: Homo sapiens
2490 DAVID IDs
Input genes

Gene name

Save results
[Help and Manual](#)

Download File

AFFYMETRIX_3PRIME_ID	Gene Name	Related Genes	Species
1007_s_at	microRNA_4640(MIR4640)	RG	Homo sapiens
1053_at	replication_factor_C_subunit_2(REC2)	RG	Homo sapiens
117_at	heat_shock_protein_family_A_(Hsp70)_member_6(HSPA6)	RG	Homo sapiens
121_at	paired_box_8(PAX8)	RG	Homo sapiens
1294_at	microRNA_5193(MIR5193)	RG	Homo sapiens
1316_at	thyroid_hormone_receptor_alpha(THRA)	RG	Homo sapiens
1431_at	cytochrome_P450_family_2_subfamily_E_member_1(CYP2E1)	RG	Homo sapiens
1487_at	estrogen_related_receptor_alpha(ESRRA)	RG	Homo sapiens
1494_f_at	cytochrome_P450_family_2_subfamily_A_member_6(CYP2A6)	RG	Homo sapiens
1598_g_at	growth_arrest_specific_6(GAS6)	RG	Homo sapiens
160020_at	matrix_metalloproteinase_14(MMP14)	RG	Homo sapiens
177_at	phospholipase_D1(PLD1)	RG	Homo sapiens
179_at	DTX2P1-UPK3BP1-PMS2P11_readthrough_transcribed_pseudogene(DTX2P1-UPK3BP1-PMS2P11)	RG	Homo sapiens
1861_at	BCL2_associated_agonist_of_cell_death(BAD)	RG	Homo sapiens
200000_s_at	pre-mRNA_processing_factor_8(PRPF8)	RG	Homo sapiens
200001_at	calpain_small_subunit_1(CAPNS1)	RG	Homo sapiens
200002_at	ribosomal_protein_L35(RPL35)	RG	Homo sapiens
200003_s_at	microRNA_5805(MIR5805)	RG	Homo sapiens
200004_at	eukaryotic_translation_initiation_factor_4_gamma_2(EIF4G2)	RG	Homo sapiens
200005_at	eukaryotic_translation_initiation_factor_3_subunit_D(EIF3D)	RG	Homo sapiens
200006_at	Parkinsonism_associated_deglycase(PARK7)	RG	Homo sapiens
200007_at	signal_recognition_particle_14(SRP14)	RG	Homo sapiens
200008_s_at	GDN_receptor_inhibitor_3(GRI3)	RG	Homo sapiens

Q1: What are the gene names of the genes with Affy_id : “1053_at” and “200010_at”?

Tool 3. Functional Annotation Tool

Go back to the previous page or choose “shortcut to DAVID Tools”—“Functional Annotation Tool”.

The screenshot shows the DAVID Analysis Wizard interface. At the top, there is a navigation bar with links: Home, Start Analysis, Shortcut to DAVID Tools, Technical Center, Downloads & APIs, Term of Service, Why DAVID?, and About Us. Below the navigation bar, a red message reads: "*** Welcome to DAVID 6.8 ***" and "*** If you are looking for DAVID 6.7, please visit our development site. ***". The main heading is "Analysis Wizard". On the left, there is a "Gene List Manager" sidebar with tabs for "Upload", "List", and "Background". The "List" tab is active, showing a dropdown menu for "Select Species" with options: "- Use All Species -", "Homo sapiens(2499)", and "Unknown(1)". Below this, there are buttons for "List Manager", "Use", "Rename", "Remove", "Combine", and "Show Gene List". The main content area shows "Step 1. Successfully submitted gene list" with "Current Gene List: List_1" and "Current Background: Homo sapiens". "Step 2. Analyze above gene list with one of DAVID tools" is followed by a blue arrow pointing to a red-bordered box containing the "Functional Annotation Tool" link. Below this box are three sub-links: "Functional Annotation Clustering", "Functional Annotation Chart", and "Functional Annotation Table". Other tool links include "Gene Functional Classification Tool", "Gene ID Conversion Tool", and "Gene Name Batch Viewer". On the right side, there are links for "Tell us how you like the tool" and "Contact us for questions", and a link for "Which DAVID tools to use?".

The functional annotation tool includes three options: Functional Annotation Clustering, chart and table. Click the “Functional Annotation Tool”, go to the new page, and choose the annotation we want (Gene Ontology and KEGG pathway for this exercise).

*** Welcome to DAVID 6.8 ***
*** If you are looking for DAVID 6.7, please visit our development site. ***

Upload List Background

Gene List Manager

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

- Use All Species -
Homo sapiens(2499)
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

[Help and Tool Manual](#)

Current Gene List: List_1 2490 DAVID IDs
Current Background: Homo sapiens Check Defaults Clear All

- Disease (0 selected)
- Functional_Categories (0 selected)
- Gene_Ontology (0 selected)
- General_Annotations (0 selected)
- Literature (0 selected)
- Main_Accessions (0 selected)
- Pathways (1 selected)

<input type="checkbox"/> BBID	3.9%	98	Chart	
<input type="checkbox"/> BIOCARTA	18.0%	449	Chart	
<input type="checkbox"/> EC_NUMBER	29.2%	727	Chart	
<input checked="" type="checkbox"/> KEGG_PATHWAY	55.1%	1372	Chart	
<input type="checkbox"/> REACTOME_PATHWAY	65.9%	1642	Chart	

- 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

Choose the KEGG pathway analysis only, and open the KEGG pathway chart.

Q2: What are the 3 most significant KEGG pathways? What are their p-values? Open them in the KEGG website.

Now choose “Functional Annotation Clustering”. The results show that pathways can be combined into 9 clusters.

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

Functional Annotation Clustering

[Help and Manual](#)

Current Gene List: List_1
 Current Background: Homo sapiens
 1889 DAVID IDs

Options Classification Stringency Medium

Overall enrichment score
The higher, the more enriched

The smaller, the more enriched.

9 Cluster(s)

Annotation Cluster	Enrichment Score		Count	P_Value	Benjamini
Annotation Cluster 1 Enrichment Score: 7.86 G ■					
<input type="checkbox"/> KEGG_PATHWAY	Parkinson's disease	RT	53	2.6E-10	2.5E-8
<input type="checkbox"/> KEGG_PATHWAY	Huntington's disease	RT	63	1.8E-9	1.3E-7
<input type="checkbox"/> KEGG_PATHWAY	Alzheimer's disease	RT	56	8.9E-9	4.3E-7
<input type="checkbox"/> KEGG_PATHWAY	Non-alcoholic fatty liver disease (NAFLD)	RT	49	2.2E-7	8.0E-6
<input type="checkbox"/> KEGG_PATHWAY	Oxidative phosphorylation	RT	44	5.6E-7	1.6E-5
Annotation Cluster 2 Enrichment Score: 1.4 G ■					
<input type="checkbox"/> KEGG_PATHWAY	Mismatch repair	RT	9	1.7E-2	1.1E-1
<input type="checkbox"/> KEGG_PATHWAY	DNA replication	RT	11	3.9E-2	1.9E-1
<input type="checkbox"/> KEGG_PATHWAY	Nucleotide excision repair	RT	12	9.4E-2	3.2E-1
Annotation Cluster 3 Enrichment Score: 0.88 G ■					
<input type="checkbox"/> KEGG_PATHWAY	Long-term potentiation	RT	19	8.6E-3	7.3E-2
<input type="checkbox"/> KEGG_PATHWAY	Oxytocin signaling pathway	RT	30	1.7E-1	4.5E-1
<input type="checkbox"/> KEGG_PATHWAY	Vascular smooth muscle contraction	RT	23	2.0E-1	4.9E-1
<input type="checkbox"/> KEGG_PATHWAY	cGMP-PKG signaling pathway	RT	30	2.5E-1	5.5E-1
<input type="checkbox"/> KEGG_PATHWAY	cAMP signaling pathway	RT	31	5.6E-1	8.1E-1
Annotation Cluster 4 Enrichment Score: 0.72 G ■					
<input type="checkbox"/> KEGG_PATHWAY	Chronic myeloid leukemia	RT	22	2.0E-3	2.3E-2
<input type="checkbox"/> KEGG_PATHWAY	Central carbon metabolism in cancer	RT	19	6.1E-3	5.4E-2
<input type="checkbox"/> KEGG_PATHWAY	Hepatitis B	RT	31	4.8E-2	2.1E-1
<input type="checkbox"/> KEGG_PATHWAY	Pancreatic cancer	RT	16	6.3E-2	2.5E-1

Q3: What do pathways have in common for annotation cluster1? What about annotation cluster 2?

Now go back and select “Functional Annotation Chart”.

The “Functional Annotation Chart” provides the clustering of genes’ annotations (KEGG pathway or others). It shows 77 chart records, which means that all the 1889 genes are included in 77 KEGG pathways.

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

Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List_1
Current Background: Homo sapiens
1889 DAVID IDs

Options

Thresholds: Count EASE

Display: Fold Enrichment Bonferroni Benjamini FDR Fisher Exact LT,PH,PT # of Records

Save files

77 chart records

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	KEGG_PATHWAY	Proteasome	RT		30	1.6	3.7E-14	1.1E-11
<input type="checkbox"/>	KEGG_PATHWAY	Ribosome	RT		53	2.8	4.0E-11	5.8E-9
<input type="checkbox"/>	KEGG_PATHWAY	Parkinson's disease	RT		53	2.8	2.6E-10	2.5E-8
<input type="checkbox"/>	KEGG_PATHWAY	Huntington's disease	RT		63	3.3	1.8E-9	1.3E-7
<input type="checkbox"/>	KEGG_PATHWAY	Biosynthesis of antibiotics	RT		67	3.5	3.0E-9	1.7E-7
<input type="checkbox"/>	KEGG_PATHWAY	Alzheimer's disease	RT		56	3.0	8.9E-9	4.3E-7
<input type="checkbox"/>	KEGG_PATHWAY	Spliceosome	RT		46	2.4	6.8E-8	2.8E-6
<input type="checkbox"/>	KEGG_PATHWAY	Non-alcoholic fatty liver disease (NAFLD)	RT		49	2.6	2.2E-7	8.0E-6
<input type="checkbox"/>	KEGG_PATHWAY	Protein processing in endoplasmic reticulum	RT		53	2.8	2.2E-7	7.2E-6
<input type="checkbox"/>	KEGG_PATHWAY	Oxidative phosphorylation	RT		44	2.3	5.6E-7	1.6E-5
<input type="checkbox"/>	KEGG_PATHWAY	Epstein-Barr virus infection	RT		56	3.0	9.1E-7	2.4E-5
<input type="checkbox"/>	KEGG_PATHWAY	Focal adhesion	RT		57	3.0	6.3E-6	1.5E-4

KEGG pathway

Related genes

Count Threshold (Minimum Count): The threshold of minimum gene counts belonging to an annotation term. Default value is 2. In short, you do not trust the term only having one gene involved.

Pathways are ordered by ascending p-value but can be ordered by any other column by clicking on the header of the column.

Q4: What is the pathway with a higher gene count?

Now choose the “Functional Annotation Table”.

The “Functional Annotation Table” shows that 1059 genes are annotated with one or more annotations (here, KEGG pathways).

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

Functional Annotation Table

[Help and Manual](#)

Current Gene List: [List_1](#)
 Current Background: [Homo sapiens](#)
 1889 DAVID IDs

Save results

1059 record(s)  [Download File](#)

203282_at	1,4-alpha-glucan branching enzyme 1(GBE1)	Related Genes	Homo sapiens
KEGG_PATHWAY	Starch and sucrose metabolism, Metabolic pathways,		
200862_at	24-dehydrocholesterol reductase(DHCR24)	Related Genes	Homo sapiens
KEGG_PATHWAY	Steroid biosynthesis, Metabolic pathways,		
203058_s_at, 203059_s_at, 203060_s_at	3'-phosphoadenosine 5'-phosphosulfate synthase 2(PAPSS2)	Related Genes	Homo sapiens
KEGG_PATHWAY	Purine metabolism, Selenocompound metabolism, Sulfur metabolism, Metabolic pathways, Biosynthesis of antibiotics,		
202539_s_at	3-hydroxy-3-methylglutaryl-CoA reductase(HMGCR)	Related Genes	Homo sapiens
KEGG_PATHWAY	Terpenoid backbone biosynthesis, Metabolic pathways, Biosynthesis of antibiotics, AMPK signaling pathway, Bile secretion,		
202772_at	3-hydroxymethyl-3-methylglutaryl-CoA lyase(HMGCL)	Related Genes	Homo sapiens
KEGG_PATHWAY	Synthesis and degradation of ketone bodies, Valine, leucine and isoleucine degradation, Butanoate metabolism, Metabolic pathways, Peroxisome,		
202419_at	3-ketodihydroshpingosine reductase(KDSR)	Related Genes	Homo sapiens
KEGG_PATHWAY	Sphingolipid metabolism, Metabolic pathways,		
202780_at	3-oxoacid CoA-transferase 1(OXCT1)	Related Genes	Homo sapiens
KEGG_PATHWAY	Synthesis and degradation of ketone bodies, Valine, leucine and isoleucine degradation, Butanoate metabolism,		
202464_s_at	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3(PFKFB3)	Related Genes	Homo sapiens
KEGG_PATHWAY	Fructose and mannose metabolism, HIF-1 signaling pathway, AMPK signaling pathway,		
202123_s_at	ABL proto-oncogene 1, non-receptor tyrosine kinase(ABL1)	Related Genes	Homo sapiens
KEGG_PATHWAY	ErbB signaling pathway, Ras signaling pathway, Cell cycle, Axon guidance, Neurotrophin signaling pathway, Pathogenic Escherichia coli infection, Shigellosis, Pathways in cancer, MicroRNAs in cancer, Chronic myeloid leukemia, Viral myocarditis,		
202604_x_at, 202603_at	ADAM metallopeptidase domain 10(ADAM10)	Related Genes	Homo sapiens
KEGG_PATHWAY	Alzheimer's disease, Epithelial cell signaling in Helicobacter pylori infection,		
200734_s_at, 200011_s_at	ADP ribosylation factor 3(ARF3)	Related Genes	Homo sapiens
KEGG_PATHWAY	Endocytosis,		
201526_at	ADP ribosylation factor 5(ARF5)	Related Genes	Homo sapiens
KEGG_PATHWAY	Endocytosis,		
203312_x_at	ADP ribosylation factor 6(ARF6)	Related Genes	Homo sapiens
KEGG_PATHWAY	Ras signaling pathway, Endocytosis, Fc gamma R-mediated phagocytosis,		
202211_at	ADP ribosylation factor GTPase activating protein 3(ARFGAP3)	Related Genes	Homo sapiens
KEGG_PATHWAY	Endocytosis,		
202956_at, 202955_s_at	ADP ribosylation factor guanine nucleotide exchange factor 1(ARFGEF1)	Related Genes	Homo sapiens
KEGG_PATHWAY	Endocytosis,		
201924_at	AF4/FMR2 family member 1(AFF1)	Related Genes	Homo sapiens

Tool 4. Gene Functional Classification

Click the “Gene Functional Classification tool” under the list of “Shortcut to DAVID tools”. The results show 106 clusters of annotations. This tool is used to cluster the functionally related genes as a group and give a score to this cluster.

The screenshot displays the Gene Functional Classification Tool interface. The top navigation bar includes links for Home, Start Analysis, Shortcut to DAVID Tools, Technical Center, Downloads & APIs, Term of Service, Why DAVID?, and About Us. A welcome message for DAVID 6.8 is shown, along with a link to the development site. The main content area is titled "Gene Functional Classification Result" and shows the current gene list as "List_1" with a background of "Homo sapiens" and 2490 DAVID IDs. The classification stringency is set to "Medium". The results are displayed in two gene groups, each with an enrichment score and a list of genes with their associated annotations.

Gene Functional Classification Result

Current Gene List: List_1
Current Background: Homo sapiens
2490 DAVID IDs

Options Classification Stringency Medium

Rerun using options Create Sublist

106 Cluster(s)

Gene Group 1		Enrichment Score: 93.06	RG	T	Map
1	<input type="checkbox"/> 203262_s_at	family with sequence similarity 50 member A(FAM50A)			
2	<input type="checkbox"/> 203023_at	NOP16 nucleolar protein(NOP16)			
3	<input type="checkbox"/> 201922_at	NSA2, ribosome biogenesis homolog(NSA2)			
4	<input type="checkbox"/> 202579_x_at	high mobility group nucleosomal binding domain 4(HMGN4)			
5	<input type="checkbox"/> 201414_s_at	nucleosome assembly protein 1 like 4(NAP1L4)			
6	<input type="checkbox"/> 200053_at	sperm associated antigen 7(SPAG7)			
7	<input type="checkbox"/> 203119_at	coiled-coil domain containing 86(CCDC86)			
8	<input type="checkbox"/> 204528_s_at	nucleosome assembly protein 1 like 1(NAP1L1)			
9	<input type="checkbox"/> 203831_at	R3H domain containing 2(R3HDM2)			
10	<input type="checkbox"/> 202882_x_at	nucleolar protein 7(NOL7)			
11	<input type="checkbox"/> 204805_s_at	H1 histone family member X(H1EX)			

Gene Group 2		Enrichment Score: 79.17	RG	T	Map
1	<input type="checkbox"/> 202791_s_at	protein phosphatase 6 regulatory subunit 2(PPP6R2)			
2	<input type="checkbox"/> 201309_x_at	neuronal regeneration related protein(NREP)			
3	<input type="checkbox"/> 201462_at	secernin 1(SCRN1)			
4	<input type="checkbox"/> 204837_at	myotubularin related protein 9(MTMR9)			
5	<input type="checkbox"/> 204793_at	G protein-coupled receptor associated sorting protein 1(GPRASP1)			

Q6: What differences can you see between gene groups?

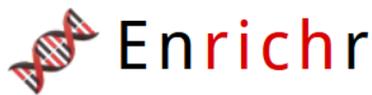
Exercise 2. Using Enrichr

Enrichr (<http://amp.pharm.mssm.edu/Enrichr/>) accepts either BED format or a list of genes with gene symbols.

1. Upload your gene list

Enrichr uses a list of gene symbols as input data. You can upload the list by either selecting the text file that contains the list or just simply pasting the list into the text box. It is better to enter a description for the gene list so that multiple lists can be differentiated from each other.

We will use the same genes from the previous exercise.



Analyze What's New? Libraries **Find a Gene** **About** Help

Input data

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

Try an example [BED file](#).

No file selected.

Text file including genes

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

```
204820_s_at
204824_at
204831_at
204832_s_at
204834_at
204835_at
204837_at
204838_s_at
204839_at
204841_s_at
```

Input gene symbols

0 gene(s) entered

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

GSE3585

Description of this dataset

Contribute

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

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;128(14).

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

2. Results page

On the results page, the analysis is divided into different categories of enrichment (Transcription, Pathways, ontologies and so on). The first category is shown. Within each category, the enrichment analyses of various gene-set libraries are listed. We open the pathway analysis as an example, presenting a multitude of visualizations. If you want to change the category, just tap the other category name.

Enrichr Login | Register

Categories

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

Description GSE3585 (7384 genes)

Category	Gene Set	Gene Count
KEGG 2016	Metabolic pathways_Homo sapiens_hsa01110	~100
	Pathways in cancer_Homo sapiens_hsa05201	~80
	Focal adhesion_Homo sapiens_hsa04510	~70
	Endocytosis_Homo sapiens_hsa04144	~60
	Epstein-Barr virus infection_Homo sapiens_hsa05100	~50
WikiPathways 2016	XPodNet - protein-protein interactions in the cytoplasm_Homo sapiens_WP411	~100
	mRNA processing_Mus musculus_WP310	~80
	PodNet: protein-protein interactions in the cytoplasm_Homo sapiens_WP411	~70
	Cytoplasmic Ribosomal Proteins_Homo sapiens_WP411	~60
	mRNA Processing_Homo sapiens_WP411	~50
ARCHS4 Kinases Coexp	YES1_human_kinase_ARCHS4_coexpression	~100
	UHMK1_human_kinase_ARCHS4_coexpression	~80
	TGFBR2_human_kinase_ARCHS4_coexpression	~70
	RYK_human_kinase_ARCHS4_coexpression	~60
	MAPK6_human_kinase_ARCHS4_coexpression	~50
Reactome 2016	Metabolism_Homo sapiens_R-HSA-1430728	~100
	Gene Expression_Homo sapiens_R-HSA-7416	~80
	Infectious disease_Homo sapiens_R-HSA-561	~70
	Disease_Homo sapiens_R-HSA-1643685	~60
	Metabolism of proteins_Homo sapiens_R-HSA-1430728	~50
BioCarta 2016	mCalpain and friends in Cell motility_Homo sapiens_Bc000001	~100
	Role of ERBB2 in Signal Transduction and Orchestration of Skeletal muscle hypertrophy is regulated via the PI3K/Akt pathway_Homo sapiens_Bc000001	~80
	Mechanism of Gene Regulation by Peroxisome Biogenesis Defect 1_Homo sapiens_Bc000001	~70
	Transcription factor CREB and its extracellular matrix signaling_Homo sapiens_Bc000001	~60
	Transcription factor CREB and its extracellular matrix signaling_Homo sapiens_Bc000001	~50
HumanCyc 2016	superpathway of conversion of glucose to acetyl-CoA_Homo sapiens_PWWY-7	~100
	protein ubiquitylation_Homo sapiens_PWWY-7	~80
	3-phosphoinositide biosynthesis_Homo sapiens_PWWY-7	~70
	TCA cycle_Homo sapiens_PWWY66-398	~60
	superpathway of inositol phosphate compounds_Homo sapiens_PWWY-7	~50
NCI-Nature 2016	PDGFR-beta signaling pathway_Homo sapiens_P000001	~100
	ErbB1 downstream signaling_Homo sapiens_P000001	~80
	Signaling events mediated by VEGFR1 and VEGFR2_Homo sapiens_P000001	~70
	mTOR signaling pathway_Homo sapiens_P000001	~60
	TGF-beta receptor signaling_Homo sapiens_P000001	~50
Panther 2016	Integrin signalling pathway_Homo sapiens_P000001	~100
	EGF receptor signaling pathway_Homo sapiens_P000001	~80
	Ubiquitin proteasome pathway_Homo sapiens_P000001	~70
	CCCR signaling map ST_Homo sapiens_P069	~60
	Angiogenesis_Homo sapiens_P000005	~50
BioPlex 2017	RRS1	~100
	SNRNP27	~80
	FGB	~70
	RPL18A	~60
	PSMB9	~50
huMAP	RPL19	~100
	RPS2	~80
	RPS18	~70
	RPL5	~60
	RPS16	~50
PPI Hub Proteins	SLC2A4	~100
	ESR1	~80
	GABARAPL1	~70
	GABARAPL2	~60
	CSNK2A1	~50
KEA 2015	CDK2	~100
	MAPK14	~80
	GSK3B	~70
	MAPK1	~60
	CDK1	~50
LINCS L1000 Kinase Perturbations down	~100	~80
	~70	~60
LINCS L1000 Kinase Perturbations up	~100	~80
	~70	~60
Kinase Perturbations from GEO down	~100	~80
	~70	~60

Click on “KEGG 2016” to view the detailed results. They include: “Bar Graph”, “Table”, “Grid”, “Network”, and “Clustergram”. When you click on the bars, you get different ranks by other score methods. Notice that it takes longer time to open “Clustergram”.

Bar Graph:

Enrichr Login | Register

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

Description GSE3585 (7384 genes) Different ways to show

KEGG 2016 Bar Graph Table Grid Network Clustergram Change color

Click the bars to sort. Now sorted by **combined score**. SVG PNG JPG

Metabolic pathways_Homo sapiens_hsa01100	Very high significance (brightest red)
Ribosome_Homo sapiens_hsa03010	High significance (dark red)
Focal adhesion_Homo sapiens_hsa04510	Medium-high significance (medium red)
Pathways in cancer_Homo sapiens_hsa05200	Medium significance (light red)
Endocytosis_Homo sapiens_hsa04144	Medium-low significance (lighter red)
Alzheimer's disease_Homo sapiens_hsa05010	Low-medium significance (light red)
Epstein-Barr virus infection_Homo sapiens_hsa05169	Low-medium significance (light red)
Non-alcoholic fatty liver disease (NAFLD)_Homo sapiens_hsa04932	Low-medium significance (light red)
Proteoglycans in cancer_Homo sapiens_hsa05205	Low-medium significance (light red)
Huntington's disease_Homo sapiens_hsa05016	Low significance (lightest red)

WikiPathways 2016

ARCHS4 Kinases Coexp

Reactome 2016

The length of the bar represents the significance of that specific gene-set or term. In addition, the brighter the color, the more significant that term is.

Table:



KEGG 2016

Bar Graph **Table** Grid Network Clustergram  

Hover each row to see the overlapping genes.

10 entries per page

Search:

Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	Metabolic pathways_Homo sapiens_hsa01100	3.221e-32	9.439e-30	-2.01	145.87
2	Ribosome_Homo sapiens_hsa03010	5.393e-28	7.901e-26	-1.72	108.11
3	Focal adhesion_Homo sapiens_hsa04510	1.167e-24	1.140e-22	-1.87	103.15
4	Alzheimer's disease_Homo sapiens_hsa05010	8.657e-22	6.341e-20	-1.77	85.69
5	Endocytosis_Homo sapiens_hsa04144	1.212e-21	7.105e-20	-1.86	89.74
6	Pathways in cancer_Homo sapiens_hsa05200	4.883e-21	2.385e-19	-1.98	92.43
7	Epstein-Barr virus infection_Homo sapiens_hsa05169	6.539e-21	2.737e-19	-1.80	83.65
8	Non-alcoholic fatty liver disease (NAFLD)_Homo sapiens_hsa05310	5.692e-20	5.692e-18	-1.88	82.04
9	Huntington disease_Homo sapiens_hsa05310	5.692e-20	5.692e-18	-1.88	76.28
10	Proteoglycan synthesis in cancer_Homo sapiens_hsa05205	5.692e-20	5.692e-18	-1.88	79.08

CBLB, FGF2, ACTB, ACTG1, IGF1R, PPP1CB, PPP1CC, CCND1, PLAU, AKT3, KDR, AKT1, PLCE1, PRKACA, PRKACB, MAP2K1, MAP2K2, PRKCB, HGF, WNT5A, RPS6, PRKCA, ANK2, ANK3, ANK1, HSPG2, TIAM1, EZR, RAF1, TP53, DDX5, SDC4, SDC2, PKN, ITPR1, TWIST1, ITPR2, PIK3R3, PIK3R2, PIK3R1, IQGAP1, HIF1A, PIK3R5, VTN, WNT6, RRAS, PLCG2, FZD1, SMAD2, FZD3, TGFB2, TGFB1, FZD5, FZD4, CAV2, FZD7, PTCH1, CAV1, FZD6, RDX, IGF2, FN1, MSN, BRAF, IGF1, ESR1, PTK2, GRB2, FGFR1, ITGB1, CDKN1A, ITGB5, ITGB3, HSPB2, PIK3CD, PIK3CB, CTSL, CASP3, TIMP3, ITGAV, RAC1, HRAS, ARHGEF12, PPP1R12A, PDPK1, MMP2, GAB1, PLAUR, RRAS2, MMP9, RHOA, DCN, MRAS, CTTN, PIK3CA, HCLS1, COL21A1, ITGA5, PPP1R12B, SOS1, TLR4, SOS2, MET, CD44, HBEGF, CAMK2B, CD63, ROCK1, ROCK2, THBS1, EGFR, CDC42, NRAS, ERBB3, ERBB4, GPC1, ERBB2, GPC3, DROSHA, FLNA, MAPK1, FLNB, FLNC, CAMK2G, EIF4B, MAPK3, LUM, STAT3, PTPN11, MAPK14, MAPK13, VEGFA, PPP1CA, RPS6KB1, PDCD4, CTNNB1, FAS, PTPN6, KRAS

When you put the cursor on the "Name", there will be a list of related genes

Download results

By clicking on the column header, you can sort the table by the term, p-value, z-score, or combined score. You can also download the table information by clicking on the "Export entries to table" button.

Grid:

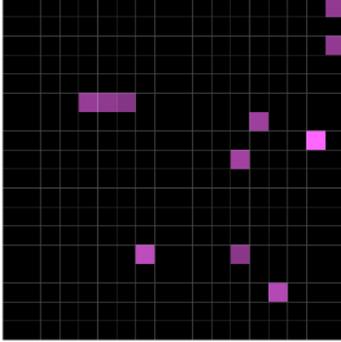
 **Enrichr** [Login](#) | [Register](#)

[Transcription](#) **[Pathways](#)** [Ontologies](#) [Disease/Drugs](#) [Cell Types](#) [Misc](#) [Legacy](#) [Crowd](#)

Description GSE3585 (7384 genes)  

KEGG 2016 [Bar Graph](#) [Table](#) **[Grid](#)** [Network](#) [Clustergram](#)  

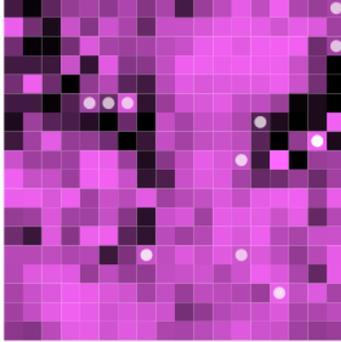
[SVG](#) [PNG](#) [JPG](#)



Z-score: -1.498
P-value: 0.06707

KEGG 2016 [Bar Graph](#) [Table](#) **[Grid](#)** [Network](#) [Clustergram](#)  

[SVG](#) [PNG](#) [JPG](#)



Z-score: -1.498
P-value: 0.06707

B cell receptor signaling pathway_Homo sapiens_hsa04662

Each grid square represents a term and is arranged based on its gene-set similarity with other terms. It shows only the top 10 terms sorted by combined score. The brighter the square, the more significant that term is. Clicking on the grid allows you to another view that colors the grid based on its correlation score with neighbors with white dots representing the significant terms. The z-score and p-value is a measure of how clustered the top 10 terms are on the grid.

Q7: Where can we find the significant gene terms?

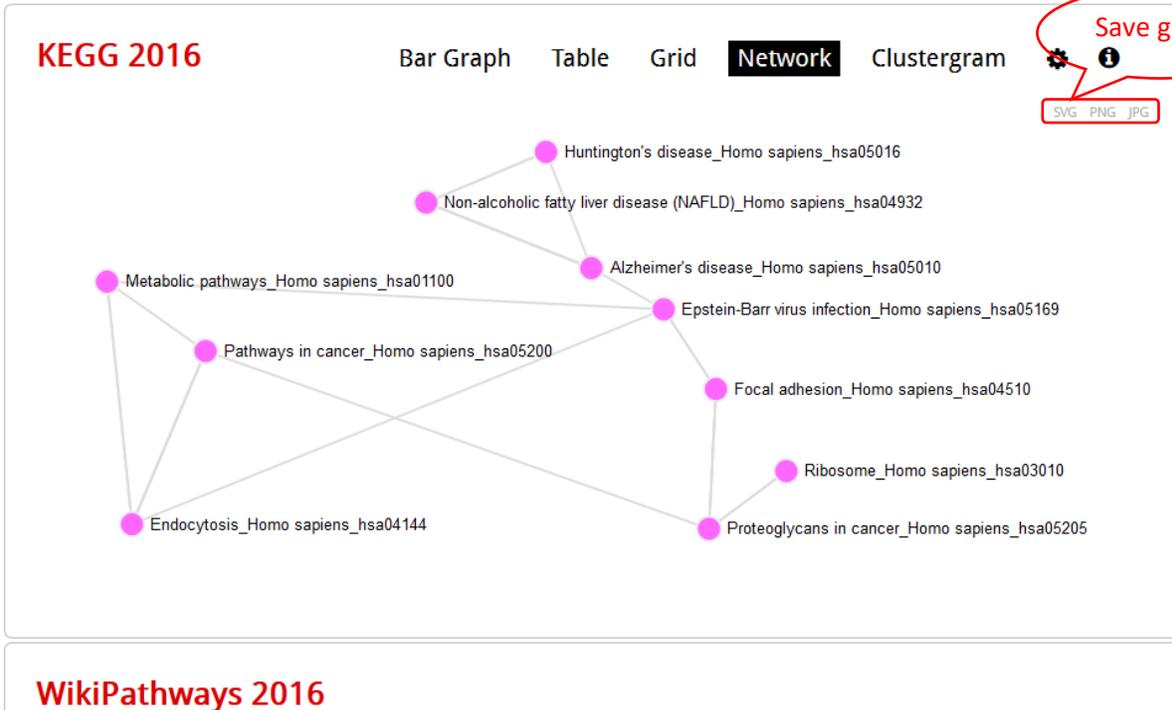
Network:



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

[Transcription](#) **[Pathways](#)** [Ontologies](#) [Disease/Drugs](#) [Cell Types](#) [Misc](#) [Legacy](#) [Crowd](#)

Description GSE3585 (7384 genes)



Each node represents a term and a link between two nodes means that the two terms have some gene content similarity.

Q8: How to find the pathway with your genes of interest?

Exercise 3. Using WebGestalt

WebGestalt(<http://www.webgestalt.org/option.php>) is a functional enrichment analysis web tool that supports three well-established and complementary methods for enrichment analysis: Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA).



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

[ORA Sample Run](#) | [GSEA Sample Run](#) | [NTA Sample Run](#) | [External Examples](#) | [Manual](#) | [Citation](#) | [User Forum](#)
[GOView](#) | [WebGestaltR](#) | [WebGestalt 2013](#)

Basic Parameters

Select Organism of Interest

Select Method of Interest

Select Functional Database

Gene List

Select Gene ID Type

| No file chosen |

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

Reference Gene List

Select Reference Set for Enrichment Analysis

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type
 | No file chosen |

Advanced parameters

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

1. Setting parameters

Set the parameters and upload the gene list, as in the following picture, and click the “Submit” button.

We are using ORA. If we change the method to “GSEA”, then we need a ranked gene list.



WEB-based Gene Set Analysis Toolkit

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

[ORA Sample Run](#) | [GSEA Sample Run](#) | [NTA Sample Run](#) | [External Examples](#) | [Manual](#) | [Citation](#) | [User Forum](#)

[GOView](#) | [WebGestaltR](#) | [WebGestalt 2013](#)

Basic Parameters

Select Organism of Interest [?]

Select Method of Interest [?]

Select Functional Database [?]

Gene List

Select Gene ID Type [?]

No file chosen

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

OR

Reference Gene List

Select Reference Set for Enrichment Analysis [?]

OR

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

No file chosen

Advanced parameters

2. Results

After we submit the task, the summary comes into being at first. It contains enrichment method, organism, enrichment category, gene list with ID type, reference gene list, and parameters for enrichment analysis. We also get: “User ID Mapping Table”, “GOSlim Summary” and “Enrichment Results”.



Summary (Result Download)

Enrich method: ORA
Organism:hsapiens
Enrichment Categories: pathway_KEGG
Interesting gene list: bioAreaUpload_1535545967.txt, ID type: affy_hg_u133a
The intersecting gene list contains 2650 user IDs in which 2325 user IDs are unambiguously mapped to the unique Entrez Gene IDs and 175 user IDs are mapped to multiple Entrez Gene IDs or could not be mapped to any Entrez Gene ID. The GO Slim summary are based upon the 2325 unique Entrez Gene IDs. Among the 2324 unique Entrez Gene IDs, 1338 IDs are annotated to the selected functional categories and also in the reference gene list, which are used for the enrichment analysis.
Reference gene list: all mapped Entrez Gene IDs from the selected platform affy_hg_u133a
The reference gene list contains 11949 IDs and 5288 IDs are annotated to the selected functional categories that are used as the reference for the enrichment analysis.
Parameters for the enrichment analysis:

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

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

2.1 User ID Mapping Table

In the table, the left contains the mapped ID, gene symbol, gene names, and Entrez gene ID. The right contains the “User IDs mapped to multiple Entrez IDs or not mapped”.



Summary User ID Mapping Table GOSlim Summary Enrichment Results

userid	Gene Symbol	Gene Name	Entrez Gene
203440_at	CDH2	cadherin 2	1000
204212_at	ACOT8	acyl-CoA thioesterase 8	10005
202382_s_at	GNPDA1	glucosamine-6-phosphate deaminase 1	10007
203415_at	PDCC6	programmed cell death 6	10018
203320_at	SH3BP3	SH3B adaptor protein 3	10019
204677_at	CDH5	cadherin 5	1003
204752_x_at	PARP2	poly(ADP-ribose) polymerase 2	10038
204485_s_at	TOML1	target of myb1 like 1 membrane trafficking protein	10040
202582_s_at	RANBP9	RAN binding protein 9	10048
201663_s_at	SMC4	structural maintenance of chromosomes 4	10051
201177_s_at	UBA2	ubiquitin like modifier activating enzyme 2	10054
203105_x_at	DNM1L	dyxnanin 1 like	10059

userid
1007_s_at
1294_at
1494_f_at
200003_s_at
200012_x_at
200026_at
200002_s_at
200033_at
200038_s_at
200047_s_at
200062_s_at
200065_s_at

2.2 GOSlim Summary

The three charts represent Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) categories, in red, blue and green bars, respectively. The height of the bar represents the number of user list genes observed in the category.

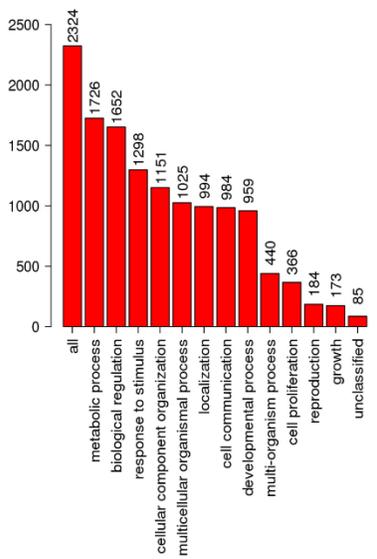


Summary | User ID Mapping Table | **GOSlim Summary** | Enrichment Results

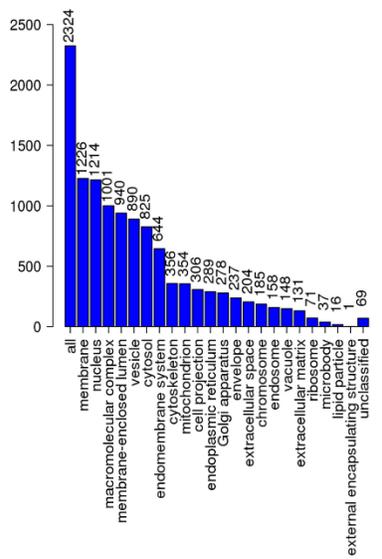
GOSlim summary for the user list genes

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

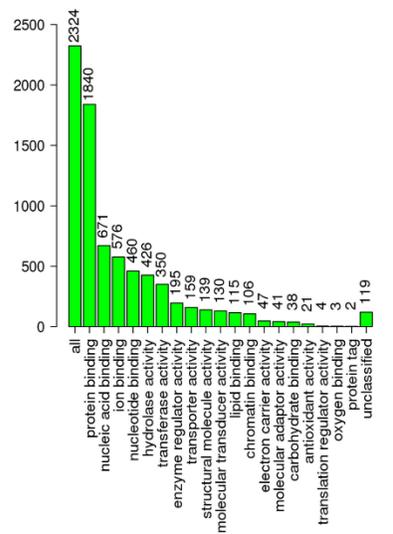
Bar chart of Biological Process categories



Bar chart of Cellular Component categories



Bar chart of Molecular Function categories



Q9: Based on the pictures, how would you describe the genes in your dataset in your own words?

2.3 Enrichment Results

The left table is the summary table, and the right one is the detailed information table.

In the right table:

“C”: the number of reference genes in the category

“O”: the number of genes in the uploaded gene list and also in the category

“E”: the expected number in the category

“R”: ratio of enrichment

“P-Value”: p-value from hypergeometric test

“FDR”: FDR from BH

WEB-based Gene Set Analysis Toolkit

Summary | User Uploading Table | GO Summary | **Enrichment Results**

Summary of the enriched categories

This table lists the enriched categories, number of enriched genes in the user gene list and also in the categories and FDR.

ID	Name	Gene	FDR
hsa03050	Proteasome - Homo sapiens (human)	30	5.22e-09
hsa05016	Huntington's disease - Homo sapiens (human)	73	5.12e-08
hsa03010	Ribosome - Homo sapiens (human)	63	7.64e-08
hsa05012	Parkinson's disease - Homo sapiens (human)	57	7.54e-08
hsa03040	Spliceosome - Homo sapiens (human)	55	1.1e-07
hsa03019	Oxidative phosphorylation - Homo sapiens (human)	51	1.17e-06
hsa05010	Alzheimer's disease - Homo sapiens (human)	66	2.99e-06
hsa05032	Non-alcoholic fatty liver disease (NAFLD) - Homo sapiens (human)	58	3.75e-05
hsa04142	Lysosome - Homo sapiens (human)	48	2.30e-04
hsa04141	Protein processing in endoplasmic reticulum - Homo sapiens (human)	58	6.49e-04

Related genes

KEGG id

pathway name

FDR

Save results

Download Table

Proteasome - Homo sapiens (human)

C=30; O=30; E=0.87; R=3.04; PValue=1.73e-11; FDR=5.22e-09

userid	Gene Symbol	Gene Name	Entrez Gene
203987_s_at	PSM3	proteasome activator subunit 3	10197
201675_s_at	PSM1	proteasome subunit alpha 1	5552
201315_s_at	PSM2	proteasome subunit alpha 2	5553
201532_s_at	PSM3	proteasome subunit alpha 3	5554
203360_s_at	PSM4	proteasome subunit alpha 4	5555
201274_s_at	PSM5	proteasome subunit alpha 5	5556
201114_s_at	PSM6	proteasome subunit alpha 6	5557
200878_s_at	PSM7	proteasome subunit alpha 7	5558
200039_s_at	PSM8	proteasome subunit beta 1	5559
200400_s_at	PSM9	proteasome subunit beta 2	5560
201400_s_at	PSM3	proteasome subunit beta 3	5561
202441_s_at	PSM4	proteasome subunit beta 4	5562
200760_s_at	PSM7	proteasome subunit beta 7	5565
204279_s_at	PSM9	proteasome subunit beta 9	5569
202509_s_at	PSM10	proteasome subunit beta 10	5599

Q10: What are the top 10 significant pathways?

Exercise 4. Compare the three websites in terms of KEGG pathways enrichment

What are the most significant pathways in each of the GSA websites?

How well do they agree?

Which website uses more databases? Which website uses more GSA methods?

Which website gives you better summary tables and figures?

What was your favorite GSA website?

