

# Functional annotation analysis: GO and KEGG Pathway analysis

**Rossella De Cegli, PhD**

(TIGEM BAD days)

Monday July 10th 2023



## DAVID is the Database for Annotation, Visualization and Integrated Discovery tool:

Provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes

The last version is the **DAVID 2021 (Dec 2021)** (<https://david.ncifcrf.gov>)

**For any given gene list, DAVID tool is able to:**

- Identify enriched biological themes (GO terms)
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathways
- Link gene-disease
- Highlight protein functional domains and motifs
- Redirect to related literatures

The session timeout of DAVID web was set to 30 min: if your web browser has no activities for 30 min, all your web session information will be flood.

David is able to efficiently annotate a list consisting of **3000 genes**.

## How to use the DAVID tool:

### I Step: Upload the list of interest

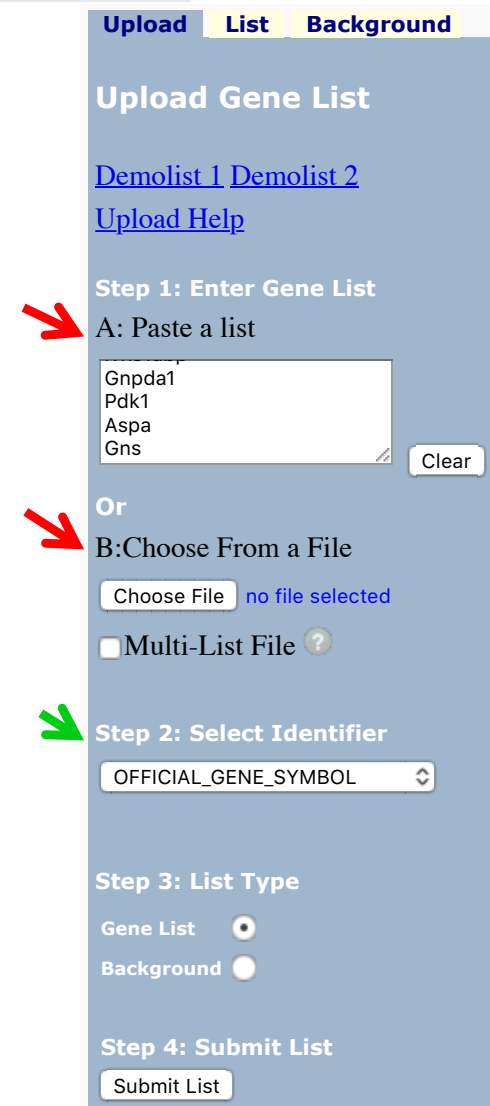
➤ You can either paste a **gene list** to the text box or load a gene list (**one gene in one row**) from a file.

➤ It is possible to Enter different list of **Identifiers**. The most used are:

- Affymetrix IDs (coming from a microarray experiment),
- Official Gene Symbols,
- RefSeq Transcript ID
- ENSEMBL gene IDs,
- ENTREZ gene IDs

### ➤ the file to UPLOAD in DAVID:

1. all the DAVID analyses must be performed on the list of induced and inhibited genes (DEGs) or proteins, **separately** (**TREATED vs UNTREATED, KO vs WT** etc..).
2. The first threshold is **FDR<5%**.
3. If the number of DEGs or proteins is too high (>1000) the additional threshold of  $\log_{2}FC > 2$ ..... $> 1.5$  for the **induced** and  $\log_{2}FC < -2$ -(.... $< -1.5$ ...it depends on the number of DEGs) for the **inhibited** ones is applied.



## II Step: Select the **Background** and the **Species**

- ✧ You may select a specific specie(s) with the List Manager on the left side of the page by highlighting the specific specie(s) and pressing the "Select" button.
- ✧ David contains information on over **1.5 million** genes from more than 65,000 species

## III Step: Select the **LEVEL** and the **output** for the analysis

- ✧ Level 1 represents the most general categories and provides the most coverage,
- ✧ Level 5 provides more specific information and less coverage.

The GO vocabulary is a type of hierarchy:

a term at level 5 is a child of a term at level 1 for a given gene.

➤ For a pivotal study use the output:

- **BP\_FAT**;
- **CC\_FAT**;
- **MF\_FAT Clustering** (or **Chart**)
- together with the **KEGG pathway analysis**



Upload List **Background**

### Gene List Manager

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

- Use All Species -

Mus musculus(275)

Rattus norvegicus(238)

Homo sapiens(235)

Select Species

List Manager [Help](#)

List\_1

List\_2

Select List to:

Use

Rename

Remove

Combine

Show Gene List

## IV Step: Select the **Statistical Options**

### Functional Annotation Clustering

[Help and Manual](#)

Current Gene List: List\_1  
Current Background: Mus musculus  
254 DAVID IDs

**Options**    Classification Stringency: Medium

<b>Kappa Similarity</b>	Similarity Term Overlap: 3	Similarity Threshold: 0.50		
<b>Classification</b>	Initial Group Membership: 3	Final Group Membership: 3	Multiple Linkage Threshold: 0.50	
<b>Enrichment Thresholds</b>	EASE: 1.0			
<b>Display</b>	<input checked="" type="checkbox"/> Fold Change	<input checked="" type="checkbox"/> Bonferroni	<input checked="" type="checkbox"/> Benjamini	<input type="checkbox"/> FDR <input type="checkbox"/> LT,PH,PT

- ✧ The threshold for statistical significance in GO: **FDR<0,1** and **Enrichment Score ≥1.5**
- ✧ The threshold for the KEGG Pathway analyses is **FDR<0,1**.
- ✧ It is then possible to Download your SIGNIFICANT results as a txt file.

Now ENJOY the DAVID RESULTS!!!!

## How to interpret the DAVID RESULTS:

Annotation Cluster 1	Enrichment Score: 11.64	Count	P-Value	Fold Change	Bonferroni	Benjamini	FDR
<input type="checkbox"/> GOTERM_BP_FAT	<a href="#">lipid metabolic process</a>	52	3.2E-14	3.3E0	1.3E-10	1.3E-10	6.0E-11
<input type="checkbox"/> GOTERM_BP_FAT	<a href="#">organic acid metabolic process</a>	43	1.2E-13	3.8E0	4.9E-10	2.4E-10	2.2E-10
<input type="checkbox"/> GOTERM_BP_FAT	<a href="#">cellular lipid metabolic process</a>	43	4.6E-13	3.6E0	1.9E-9	6.4E-10	8.6E-10
<input type="checkbox"/> GOTERM_BP_FAT	<a href="#">monocarboxylic acid metabolic process</a>	32	5.1E-12	4.5E0	2.1E-8	5.3E-9	9.6E-9
<input type="checkbox"/> GOTERM_BP_FAT	<a href="#">carboxylic acid metabolic process</a>	38	8.0E-12	3.7E0	3.3E-8	6.7E-9	1.5E-8
<input type="checkbox"/> GOTERM_BP_FAT	<a href="#">oxoacid metabolic process</a>	38	1.0E-11	3.7E0	4.2E-8	7.0E-9	1.9E-8
<input type="checkbox"/> GOTERM_BP_FAT	<a href="#">fatty acid metabolic process</a>	24	5.0E-10	5.0E0	2.1E-6	3.0E-7	9.3E-7

Number of **unique DAVID gene IDs** corresponding to the input gene list.  
NB: If two or more of your identifiers represent alternatively spliced forms of the same gene it will only be counted once and reflected in the histograms

GO Terms refer to the BPs in which the genes of your list are **clusterized and mainly functionally involved**

Functional Related BP Terms

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

OFFICIAL_GENE_SYMBOL	GENE_NAME	Related Genes	Species
Cd36	CD36 antigen (Cd36)	RG	Mus musculus
Cd74	CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated)(Cd74)	RG	Mus musculus
Cd81	CD81 antigen (Cd81)	RG	Mus musculus
St6galnac2	ST6 (alpha-N-acetylneuraminyl-2,3-beta-galactosyl-1,3)-N-acetylglactosaminide alpha-2,6-sialyltransferase 2 (St6galnac2)	RG	Mus musculus
Aacs	acetoacetyl-CoA synthetase (Aacs)	RG	Mus musculus
Acacb	acetyl-Coenzyme A carboxylase beta (Acacb)	RG	Mus musculus
Acot1	acyl-CoA thioesterase 1 (Acot1)	RG	Mus musculus
Acot2	acyl-CoA thioesterase 2 (Acot2)	RG	Mus musculus
Acot3	acyl-CoA thioesterase 3 (Acot3)	RG	Mus musculus
Aldh3a2	aldehyde dehydrogenase family 3, subfamily A2 (Aldh3a2)	RG	Mus musculus
Apoa4	apolipoprotein A-IV (Apoa4)	RG	Mus musculus
Bmp2	bone morphogenetic protein 2 (Bmp2)	RG	Mus musculus
Crat	carnitine acetyltransferase (Crat)	RG	Mus musculus

## RESULTS KEGG:

Category	Term	Count	Genes	FDR
KEGG_PATHWAY	rno04141:Protein processing in endoplasmic reticulum	13	HSP90AB1, HSPBP1, HSP90B1, TUSC3, RRBP1, CKAP4, UBE2G1, HSPA4L, SIL1, CALR, PRKCSH, SEC63, UBE2E2	0,22092472
KEGG_PATHWAY	rno05012:Parkinson's disease	10	SEPT5, UQCR10, SLC25A4, GNAI2, UBE2G1, UCHL1, COX4I2, COX6A1, COX4I1, ATP5G1	5,11736064
KEGG_PATHWAY	rno04512:ECM-receptor interaction	7	LAMA4, COL27A1, COL6A1, COL2A1, SV2A, THBS1, COL5A2	12,0028716
KEGG_PATHWAY	rno04151:PI3K-Akt signaling pathway	15	HSP90AB1, FGFR1, FGFR3, COL2A1, COL5A2, VEGFC, HSP90B1, LAMA4, CCND1, GSK3B, COL27A1, COL6A1, EFNA4, THBS1, PPP2R3C	12,2111253
KEGG_PATHWAY	rno03050:Proteasome	5	PSMB10, PSMB6, PSMC5, PSME1, PSMC1	18,1856134
KEGG_PATHWAY	rno04260:Cardiac muscle contraction	6	ATP1B1, UQCR10, COX4I2, COX6A1, COX4I1, TPM2	24,8390842
KEGG_PATHWAY	rno05205:Proteoglycans in cancer	10	ACTB, FZD9, FGFR1, CCND1, CBL, CAMK2B, MSN, THBS1, TIMP3, PTPN11	25,7314002
KEGG_PATHWAY	rno05100:Bacterial invasion of epithelial cells	6	ACTB, DNM3, ARPC3, CBL, SEPT8, SEPT9	28,1532162
KEGG_PATHWAY	rno05206:MicroRNAs in cancer	8	GLS2, CCND1, FGFR3, ZEB2, MARCKS, THBS1, TIMP3, CDC25A	30,5943478
KEGG_PATHWAY	rno00190:Oxidative phosphorylation	8	UQCR10, ATP5J2, COX4I2, COX6A1, COX4I1, ATP5G1, ATP5I, PPA1	31,4514347
KEGG_PATHWAY	rno04510:Focal adhesion	10	ACTB, VEGFC, CCND1, LAMA4, GSK3B, COL27A1, COL6A1, COL2A1, THBS1, COL5A2	31,7547962
KEGG_PATHWAY	rno04974:Protein digestion and absorption	6	SLC1A5, ATP1B1, COL27A1, COL6A1, COL2A1, COL5A2	37,6954593
KEGG_PATHWAY	rno05230:Central carbon metabolism in cancer	5	GLS2, SLC16A3, SLC1A5, FGFR1, FGFR3	42,5997807
KEGG_PATHWAY	rno04120:Ubiquitin mediated proteolysis	7	FBXW8, WWP2, UBR5, UBE2G1, CBL, CDC34, UBE2E2	59,919316
KEGG_PATHWAY	rno00071:Fatty acid degradation	4	ACSL1, ACADS, ALDH2, ACADL	61,5283637
KEGG_PATHWAY	rno04961:Endocrine and other factor-regulated calcium reabsorption	4	VDR, DNM3, ATP1B1, AP2S1	61,5283637