Functional Annotation tools:

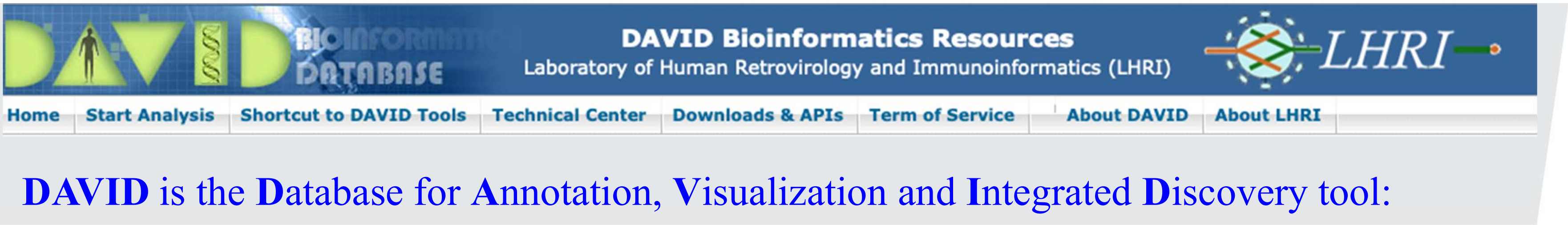
- ✧ Gene Ontology analysis
- ✧ KEGG Pathway analysis

Rossella De Cegli, PhD

(BAD days)

11/07/2022

FONDAZIONE
Telethon



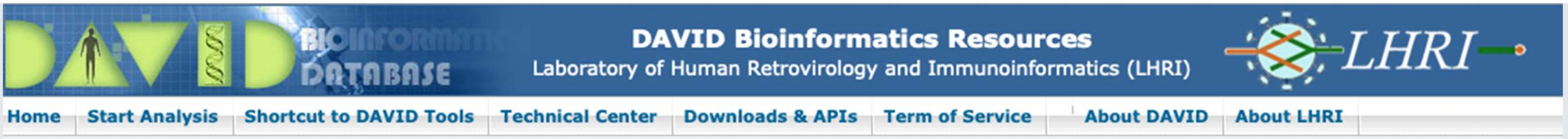
DAVID Bioinformatics Resources
Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

DAVID is the **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery 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 v6.8 (https://david.ncifcrf.gov);** Last updated June 2022

- 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**.

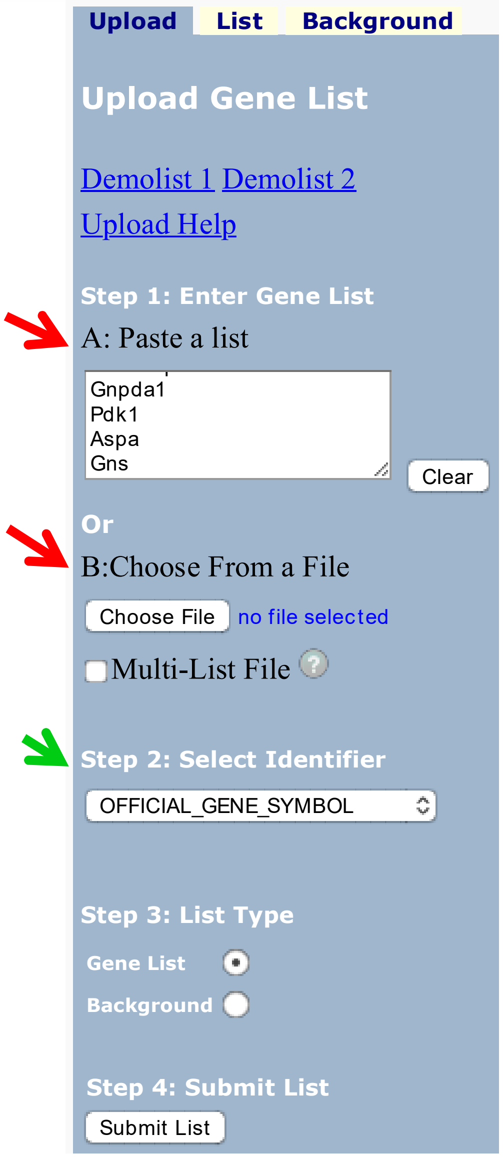

DAVID Bioinformatics Resources
Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

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,
  - Official Gene Symbols,
  - RefSeq Transcript ID
  - ENSEMBL gene IDs,
  - ENTREZ gene IDs

Exercise upload Template_David Functional list.


Upload Gene List

Step 1: Enter Gene List
A: Paste a list

```
Gnpda1
Pdk1
Aspa
Gns
```

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

Step 2: Select Identifier
OFFICIAL_GENE_SYMBOL

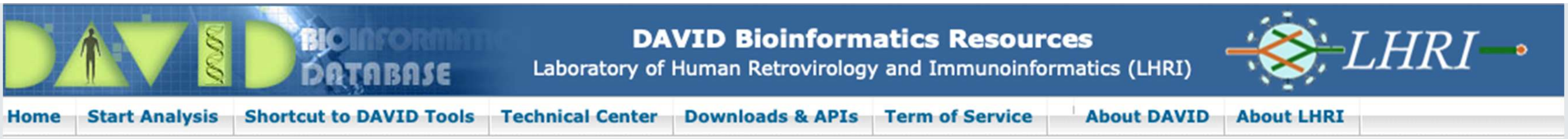
Step 3: List Type
Gene List ●
Background ○

Step 4: Submit List

HOW can we prepare 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%**. If the number of DEGs or proteins is too high (>1000) the additional threshold of logFC >2…..>1.5 for the induced and logFC<-2-(….<-1.5…it depens from the number of DEGs) for the inhibited ones is applied.

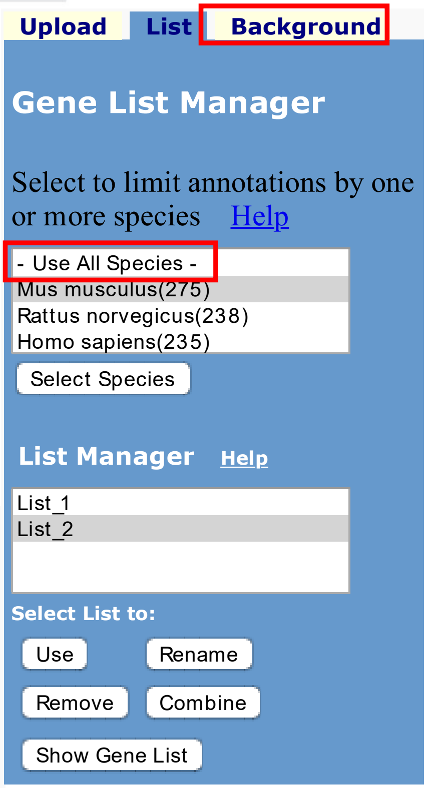
Refer to EXP1vsCTR1.xlsx



How to use the **DAVID** tool:

II Step: Selection of the Background and the Species

- ✧ Selection of the specie(s) with the List Manager on the left side of the page
- ✧ David contains information on over **1.5 million** genes from more than 65,000 species



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)

List Manager Help

List_1
List_2

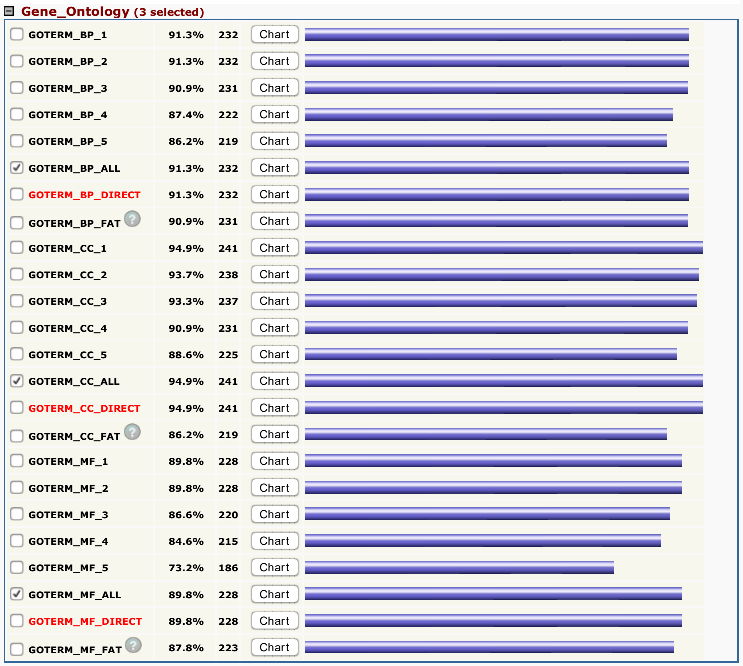
Select List to:

III Step: Selection of 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.

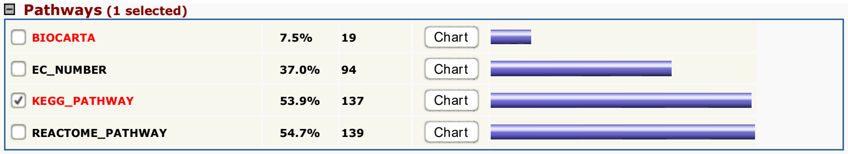


Gene_Ontology (3 selected)

☐	GOTERM_BP_1	91.3%	232	Chart
☐	GOTERM_BP_2	91.3%	232	Chart
☐	GOTERM_BP_3	90.9%	231	Chart
☐	GOTERM_BP_4	87.4%	222	Chart
☐	GOTERM_BP_5	86.2%	219	Chart
☑	GOTERM_BP_ALL	91.3%	232	Chart
☐	GOTERM_BP_DIRECT	91.3%	232	Chart
☐	GOTERM_BP_FAT	90.9%	231	Chart
☐	GOTERM_CC_1	94.9%	241	Chart
☐	GOTERM_CC_2	93.7%	238	Chart
☐	GOTERM_CC_3	93.3%	237	Chart
☐	GOTERM_CC_4	90.9%	231	Chart
☐	GOTERM_CC_5	88.6%	225	Chart
☑	GOTERM_CC_ALL	94.9%	241	Chart
☐	GOTERM_CC_DIRECT	94.9%	241	Chart
☐	GOTERM_CC_FAT	86.2%	219	Chart
☐	GOTERM_MF_1	89.8%	228	Chart
☐	GOTERM_MF_2	89.8%	228	Chart
☐	GOTERM_MF_3	86.6%	220	Chart
☐	GOTERM_MF_4	84.6%	215	Chart
☐	GOTERM_MF_5	73.2%	186	Chart
☑	GOTERM_MF_ALL	89.8%	228	Chart
☐	GOTERM_MF_DIRECT	89.8%	228	Chart
☐	GOTERM_MF_FAT	87.8%	223	Chart

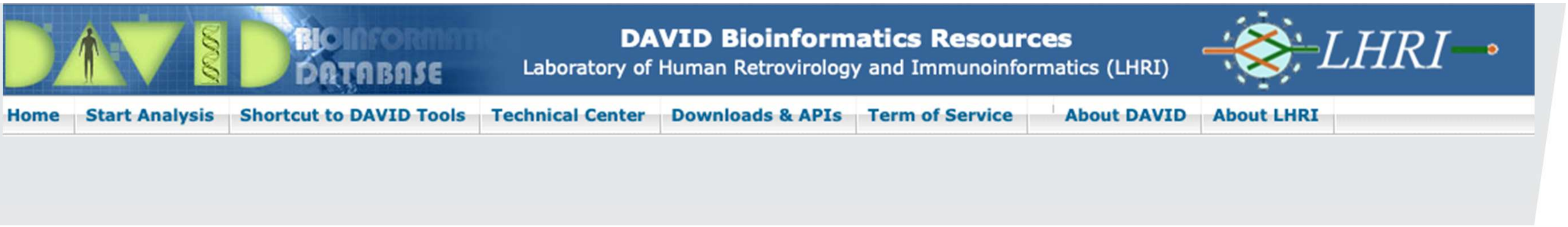
➢For a pivotal study use the output:

- ➢ **BP_FAT;**
- ➢ **CC_FAT;**
- ➢ **MF_FAT** Clustering (or Chart)

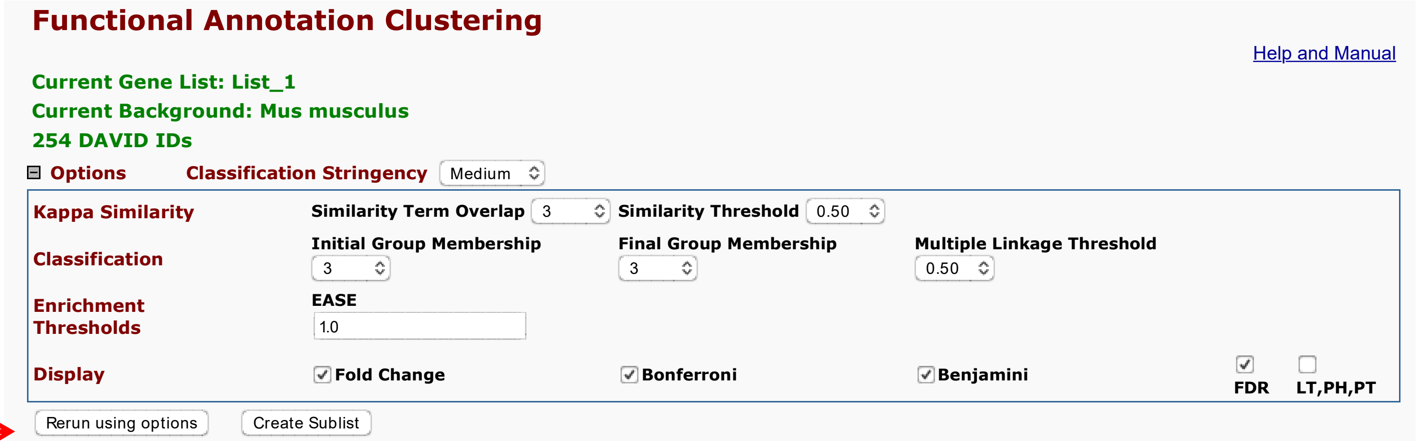


Pathways (1 selected)

☐	BIOCARTA	7.5%	19	Chart
☐	EC_NUMBER	37.0%	94	Chart
☑	KEGG_PATHWAY	53.9%	137	Chart
☐	REACTOME_PATHWAY	54.7%	139	Chart



IV Step: Selection of the Statistical Options



Functional Annotation Clustering

Help and Manual

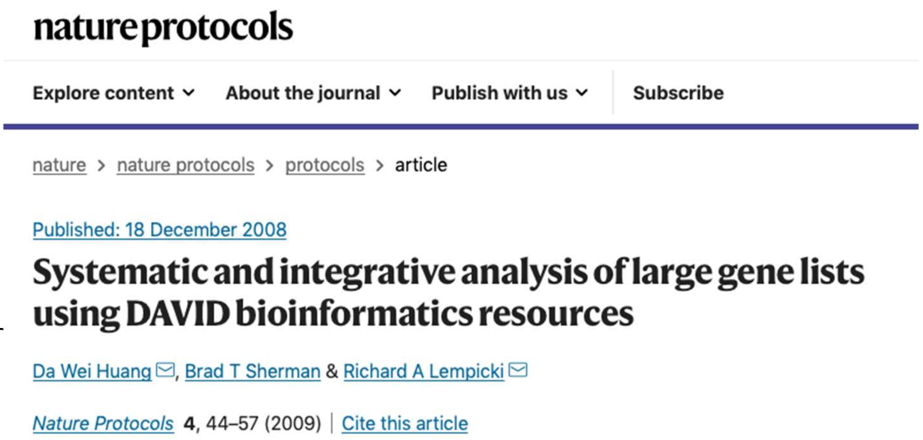
Current Gene List: List_1
Current Background: Mus musculus
254 DAVID IDs

⊟ **Options** **Classification Stringency** Medium

Kappa Similarity	**Similarity Term Overlap** 3	**Similarity Threshold** 0.50			
Classification	**Initial Group Membership** 3	**Final Group Membership** 3	**Multiple Linkage Threshold** 0.50		
Enrichment Thresholds	**EASE** 1.0				
Display	☑ **Fold Change**	☑ **Bonferroni**	☑ **Benjamini**	☑ **FDR**	☐ **LT,PH,PT**

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

Now ENJOY the DAVID RESULTS!!!!!



nature › nature protocols › protocols › article

Published: 18 December 2008

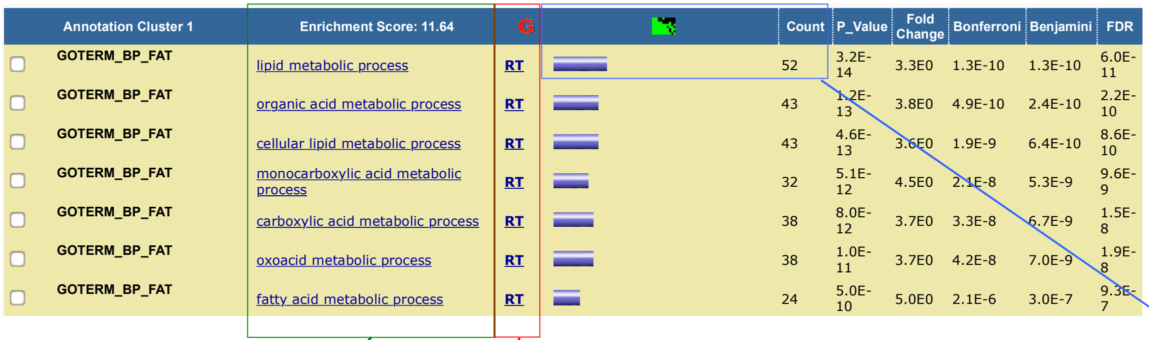
Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources

Da Wei Huang, Brad T Sherman & Richard A Lempicki

Nature Protocols **4**, 44–57 (2009) | Cite this article

visualize the results in BP_output.txt

How to interpret the DAVID RESULTS:

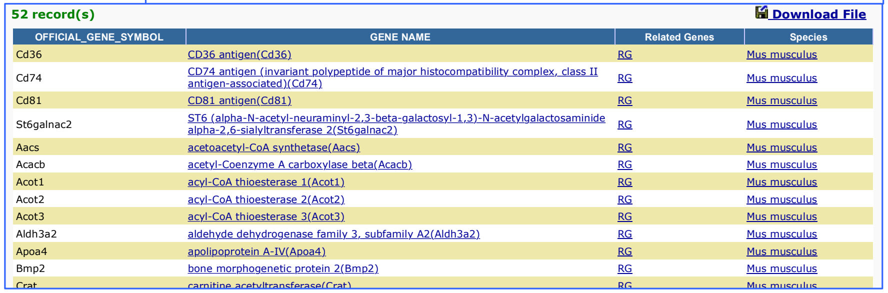


	Annotation Cluster 1	Enrichment Score: 11.64	G		Count	P_Value	Fold Change	Bonferroni	Benjamini	FDR
☐	GOTERM_BP_FAT	lipid metabolic process	RT		52	3.2E-14	3.3E0	1.3E-10	1.3E-10	6.0E-11
☐	GOTERM_BP_FAT	organic acid metabolic process	RT		43	1.2E-13	3.8E0	4.9E-10	2.4E-10	2.2E-10
☐	GOTERM_BP_FAT	cellular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

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

Functional Related BP Terms

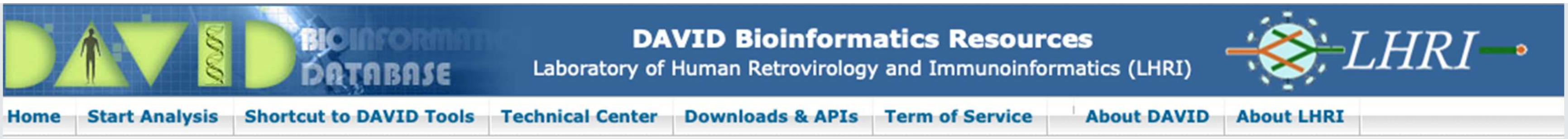
Number of **unique DAVID gene IDs** corresponding to the input gene list. IF two or more of your identifiers represent alternatively spliced forms of the same gene it will only be counted once and reflected in the histograms



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-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide 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

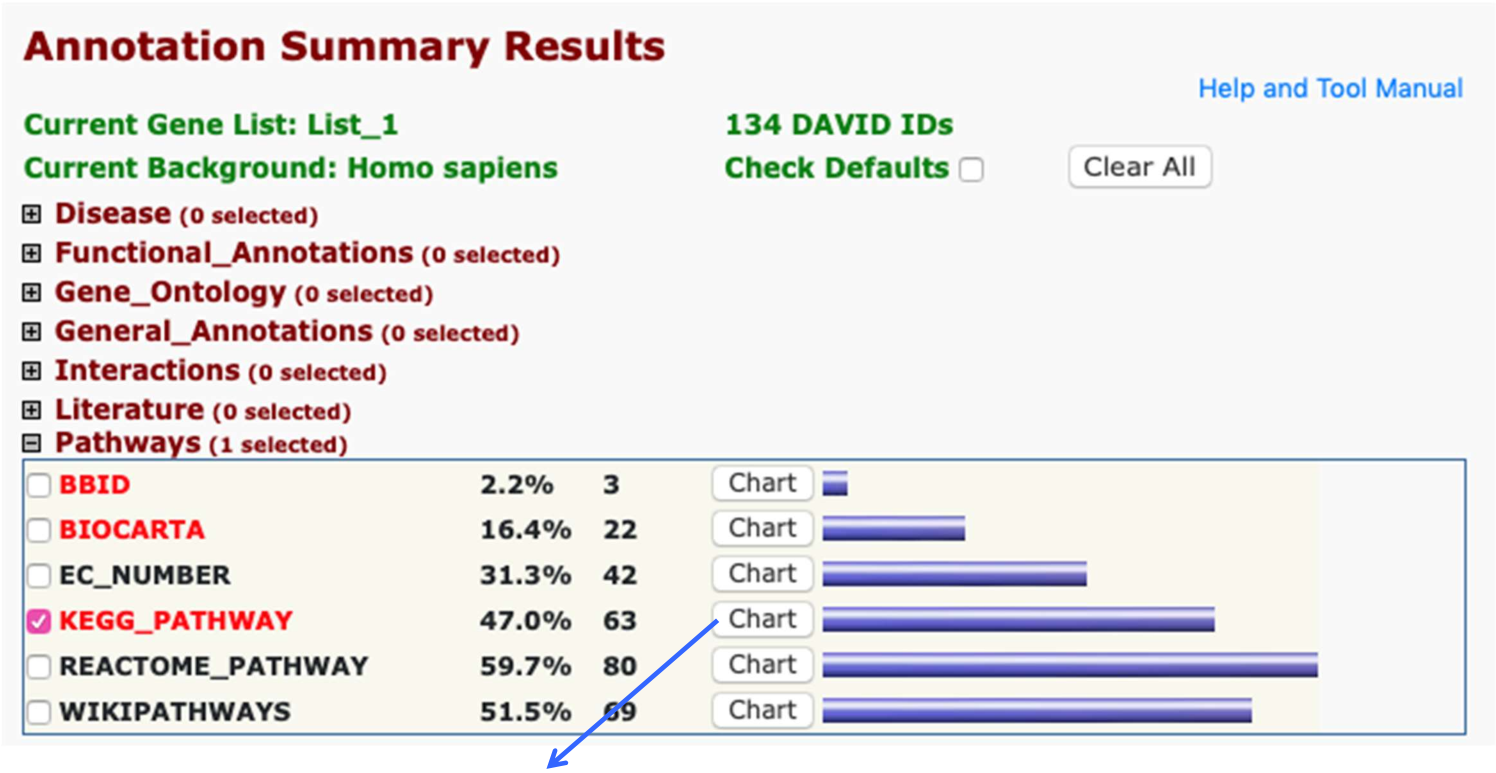

DAVID Bioinformatics Resources

Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

KEGG Pathway

KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions and relations


Annotation Summary Results

Help and Tool Manual

Current Gene List: List_1 — 134 DAVID IDs

Current Background: Homo sapiens — Check Defaults ☐ — Clear All

⊞ Disease (0 selected)

⊞ Functional_Annotations (0 selected)

⊞ Gene_Ontology (0 selected)

⊞ General_Annotations (0 selected)

⊞ Interactions (0 selected)

⊞ Literature (0 selected)

⊟ Pathways (1 selected)

☐ BBID	2.2%	3	Chart	
☐ BIOCARTA	16.4%	22	Chart	
☐ EC_NUMBER	31.3%	42	Chart	
☑ KEGG_PATHWAY	47.0%	63	Chart	
☐ REACTOME_PATHWAY	59.7%	80	Chart	
☐ WIKIPATHWAYS	51.5%	69	Chart	

- ✧ The threshold for the KEGG Pathway analyses is **FDR<0,1**.
- ✧ It is then possible to Download your SIGNIFICANT results as a txt file.

---

visualize the KEGG results in the file KEGG_output.xlsx