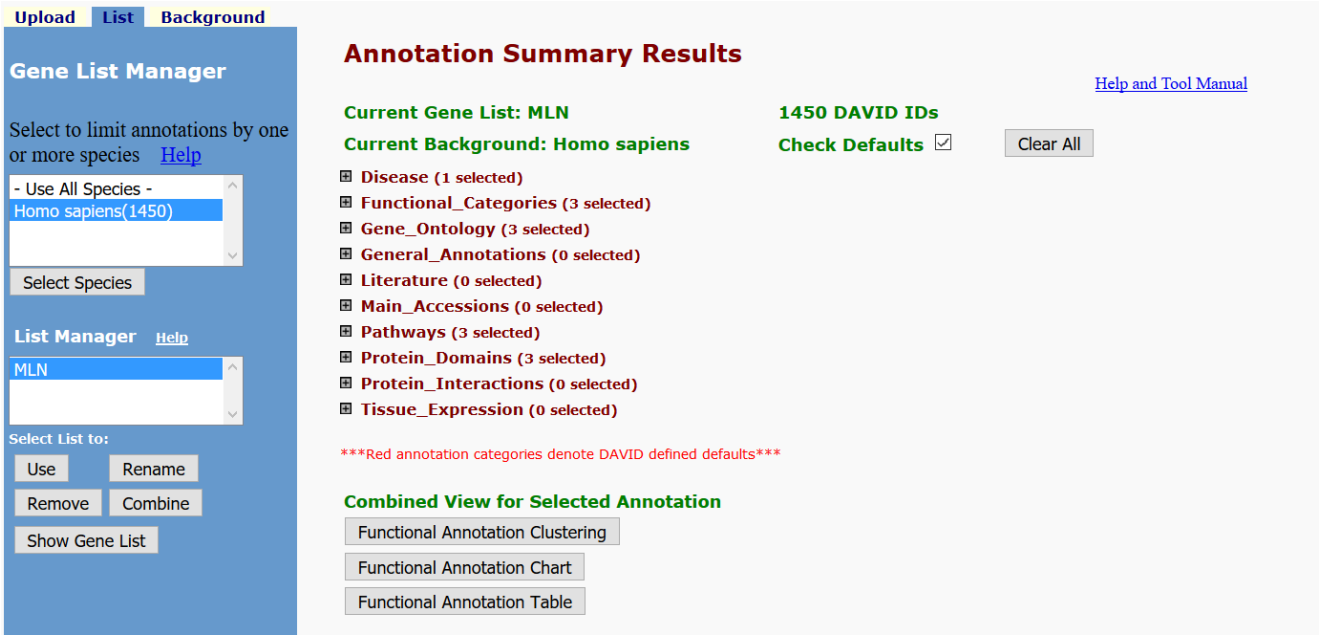
An example on using DAVID for Gene Set Enrichment and KEGG Pathway Analysis

We used DAVID for gene set enrichment analysis. Using Entrez gene IDs of differentially expressed genes we identified with our Top Tables from linear model fitting, we submitted 7 lists (by tissue type) to the DAVID database and selected the Identifier as “Entrez_GENE_ID”.

From this, we produce a list of hits that DAVID recognizes using *Homo sapiens* as a background; even though we looked at rhesus monkeys, we wanted to link biological function to humans.

In this case, for the **mesenteric lymph node**, 1450 genes out of the 1489 differentially expressed genes were recognized by DAVID. This makes sense as some probes may be redundant for the same gene.



In summary of the workflow for DAVID, we loaded the gene list -> Viewed summary page -> Explored Functional Annotation Clustering Report and selected Pathways -> KEGG_Pathway as the annotation source of interest.

DAVID uses a modified Fisher’s Exact Test to measure gene enrichment in the annotation terms; they call this EASE Scores, thus testing the significance of the association between two kinds of classification. Therefore, it would be great to explore further, out of the differentially expressed genes, which pathways are enriched. To reduce the redundancy of the functional annotation reports which dilutes biological focus, it was proposed that by grouping similar annotations together, it presents a clearer focus on the biology. As such, the grouping algorithm assumes that similar gene annotations have similar gene members.

The output for the Functional Annotation Clustering Report shows as such:

Functional Annotation Clustering

Help and Manual

Current Gene List: MLN
Current Background: Homo sapiens
1450 DAVID IDs

⊞ Options Classification Stringency: Medium

Rerun using options | Create Sublist

162 Cluster(s) Download File

	Annotation Cluster 1	Enrichment Score: 5.21				Count	P_Value	Benjamini
	GOTERM_BP_DIRECT	SRP-dependent cotranslational protein targeting to membrane	RT			24	4.9E-8	1.8E-4
	GOTERM_BP_DIRECT	translational initiation	RT			29	1.2E-7	2.2E-4
	UP_KEYWORDS	Ribosomal protein	RT			33	2.4E-7	3.7E-5
	GOTERM_BP_DIRECT	viral transcription	RT			25	3.7E-7	4.6E-4
	KEGG_PATHWAY	Ribosome	RT			28	7.9E-7	2.1E-4
	GOTERM_CC_DIRECT	ribosome	RT			30	1.4E-6	8.7E-4
	GOTERM_MF_DIRECT	structural constituent of ribosome	RT			36	2.2E-6	2.6E-3
	GOTERM_BP_DIRECT	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	RT			24	4.2E-6	3.9E-3
	UP_KEYWORDS	Ribonucleoprotein	RT			40	1.4E-5	1.3E-3
	GOTERM_BP_DIRECT	translation	RT			37	1.8E-5	1.4E-2
	GOTERM_BP_DIRECT	rRNA processing	RT			31	1.2E-4	7.1E-2
	GOTERM_CC_DIRECT	cytosolic large ribosomal subunit	RT			14	4.5E-4	9.1E-2
	GOTERM_CC_DIRECT	cytosolic small ribosomal subunit	RT			11	1.0E-3	1.2E-1
	GOTERM_CC_DIRECT	small ribosomal subunit	RT			8	1.5E-3	1.5E-1
	Annotation Cluster 2	Enrichment Score: 1.89				Count	P_Value	Benjamini
	GOTERM_BP_DIRECT	protein polyubiquitination	RT			25	1.5E-3	4.7E-1
	UP_KEYWORDS	Ligase	RT			36	5.1E-3	2.0E-1
	INTERPRO	Zinc finger, RING/FYVE/PHD-type	RT			45	7.3E-3	1.0E0
	INTERPRO	Zinc finger, RING-type	RT			32	7.6E-3	1.0E0
	GOTERM_MF_DIRECT	ubiquitin-protein transferase activity	RT			35	8.8E-3	9.3E-1
	GOTERM_MF_DIRECT	ligase activity	RT			30	8.9E-3	8.8E-1
	GOTERM_MF_DIRECT	ubiquitin protein ligase activity	RT			22	1.5E-2	8.9E-1
	SMART	RING	RT			25	2.2E-2	9.8E-1
	UP_SEQ_FEATURE	zinc finger region:RING-type	RT			24	2.6E-2	1.0E0
	UP_KEYWORDS	Ubl conjugation pathway	RT			57	3.1E-2	5.1E-1
	GOTERM_BP_DIRECT	protein ubiquitination	RT			34	4.5E-2	9.9E-1
	INTERPRO	Zinc finger, RING-type, conserved site	RT			18	5.7E-2	1.0E0
	Annotation Cluster 3	Enrichment Score: 1.47				Count	P_Value	Benjamini
	GOTERM_MF_DIRECT	potassium channel regulator activity	RT			8	1.6E-2	8.9E-1

Based on the above, DAVID provides a Group Enrichment Score, which is the geometric mean (-log scale) of p-values within the annotation cluster, to rank biological significance. It seems for this gene list, ribosome-related gene function is the most enriched. We then wanted to verify this with pathway analysis.

To confirm this, we looked specifically at the KEGG_Pathway. 493 genes from our gene list were involved in KEGG.

Functional Annotation Chart

Help and Manual

Current Gene List: List_1
Current Background: Homo sapiens
1450 DAVID IDs

⊞ **Options**

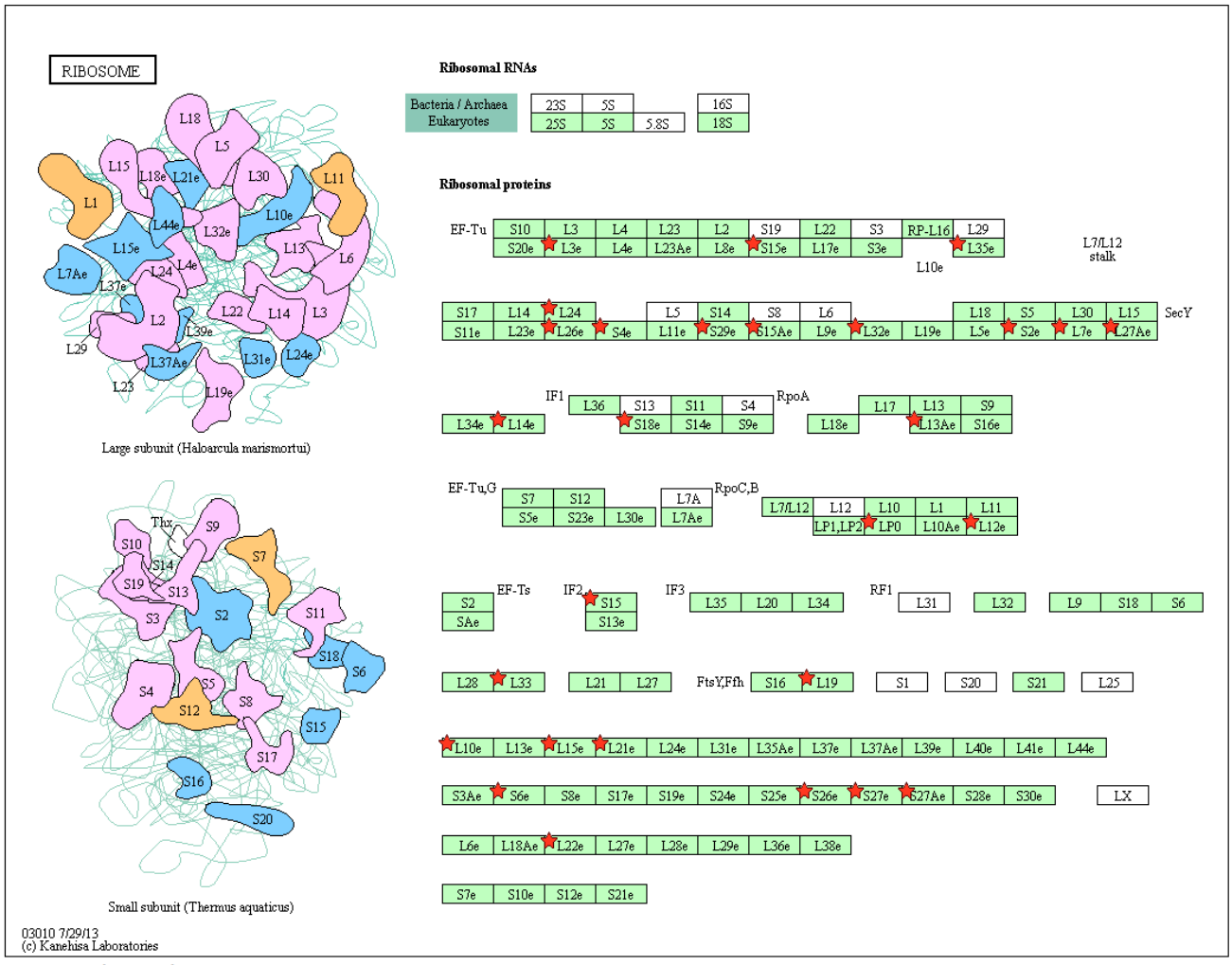
Rerun Using Options | Create Sublist

4 chart records — Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
☐	KEGG_PATHWAY	Ribosome	RT		28	1.9	7.9E-7	2.1E-4
☐	KEGG_PATHWAY	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	RT		12	0.8	1.1E-2	7.7E-1
☐	KEGG_PATHWAY	Cholinergic synapse	RT		14	1.0	5.1E-2	9.9E-1
☐	KEGG_PATHWAY	Ubiquitin mediated proteolysis	RT		16	1.1	6.2E-2	9.9E-1

1380 gene(s) from your list are not in the output.

KEGG suggests that genes relating to the ribosome are the most significant by comparison.



Following list genes shown in red above
mitochondrial ribosomal protein L19(MRPL19)
mitochondrial ribosomal protein L24(MRPL24)
mitochondrial ribosomal protein L33(MRPL33)
mitochondrial ribosomal protein S15(MRPS15)
ribosomal protein L10(RPL10)
ribosomal protein L12(RPL12)
ribosomal protein L13a(RPL13A)
ribosomal protein L14(RPL14)
ribosomal protein L15(RPL15)
ribosomal protein L21(RPL21)
ribosomal protein L22(RPL22)
ribosomal protein L26(RPL26)
ribosomal protein L27a(RPL27A)
ribosomal protein L3(RPL3)
ribosomal protein L32(RPL32)
ribosomal protein L35(RPL35)
ribosomal protein L7(RPL7)
ribosomal protein S15(RPS15)
ribosomal protein S15a(RPS15A)
ribosomal protein S18(RPS18)
ribosomal protein S2(RPS2)
ribosomal protein S26(RPS26)
ribosomal protein S27 like(RPS27L)
ribosomal protein S27a(RPS27A)
ribosomal protein S29(RPS29)
ribosomal protein S4, X-linked(RPS4X)
ribosomal protein S6(RPS6)
ribosomal protein lateral stalk subunit P0(RPLP0)

The KEGG pathway map is automatically produced by the KEGG Pathway Database. This provides a visual interpretation of the possible molecular pathway for our genes. The green colouring of the boxes represents organism-specific pathway and white boxes mean they are from the reference pathway. Although arrows are not present from the image, which would indicate molecular interaction, the boxes that connect together suggest a protein-protein interaction. We witnessed similar results in other tissues where gene enrichment analysis showed that ribosomal genes were most enriched for all differentially expressed genes by tissue type. We used another method from R as well to confirm KEGG pathway analysis