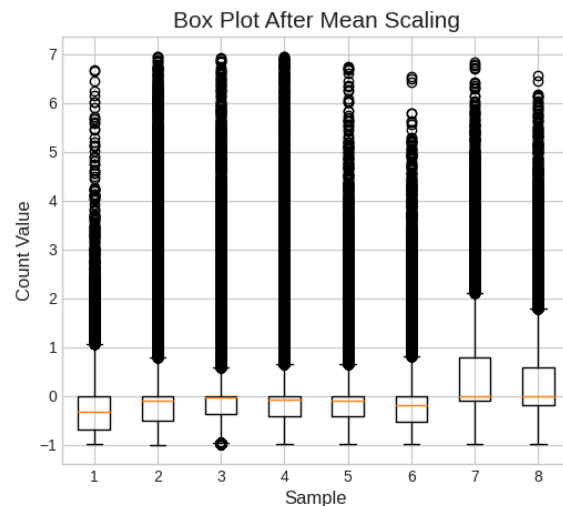
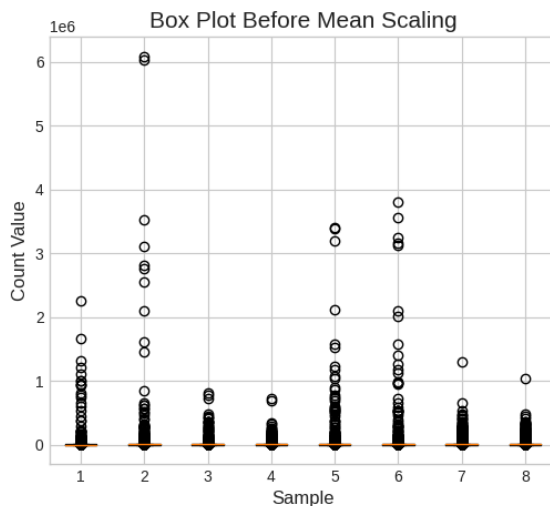


Computational Biology Lab

Assignment 3 – BE21B037

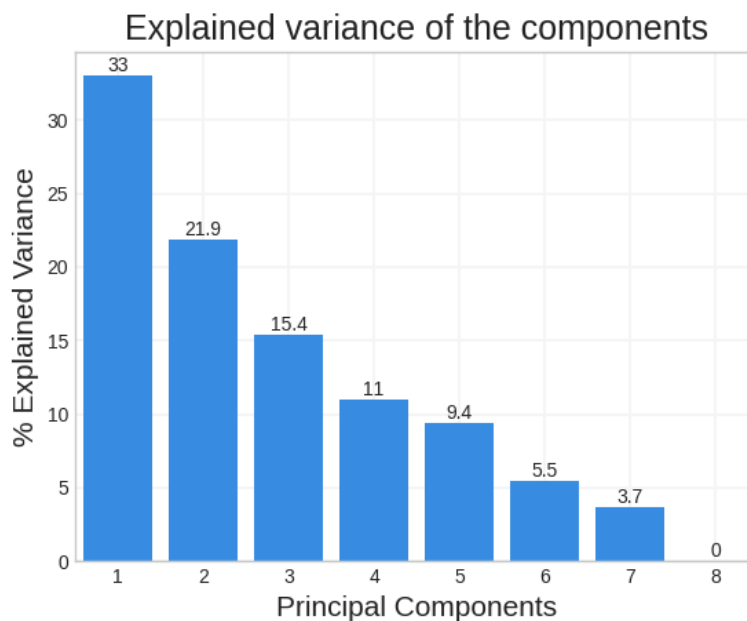
1. Principal Component Analysis (PCA):

- Apply Principal Component Analysis (PCA) to this dataset to reduce the dimensionality of the gene expression data.
- Ensure the data is scaled so that each gene (row) has a mean of zero before performing PCA.



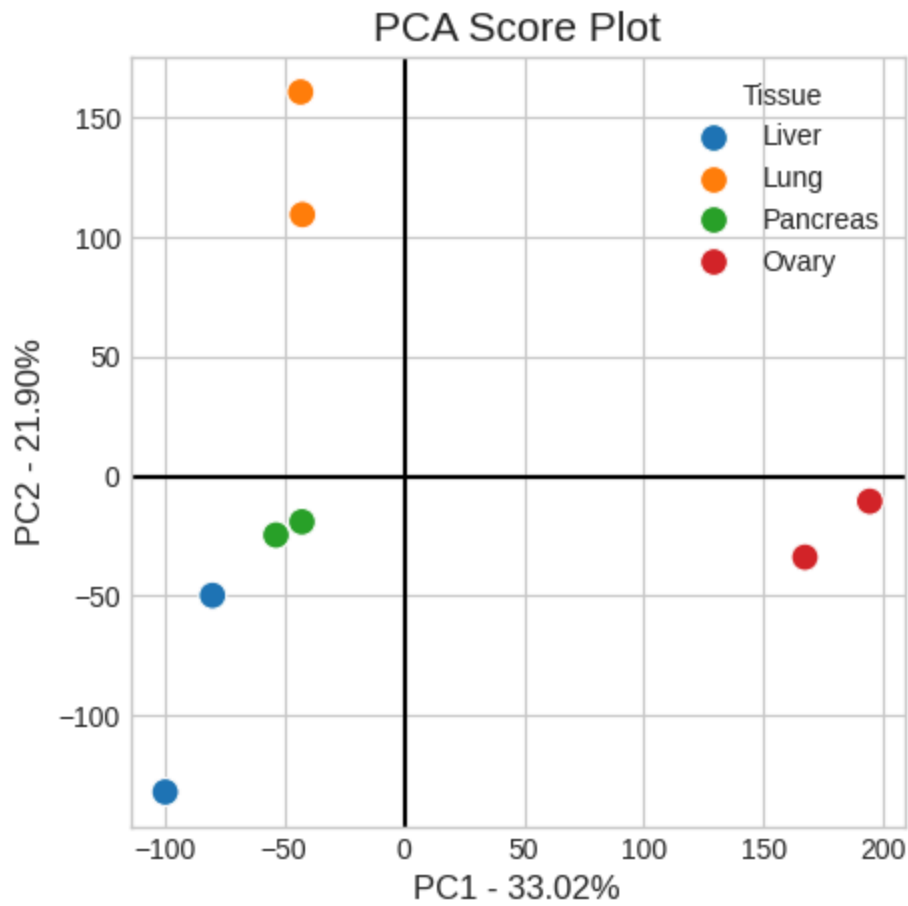
Box plot before and after scaling is provided above

- Provide the percentage of variance explained by the 8 principal components.



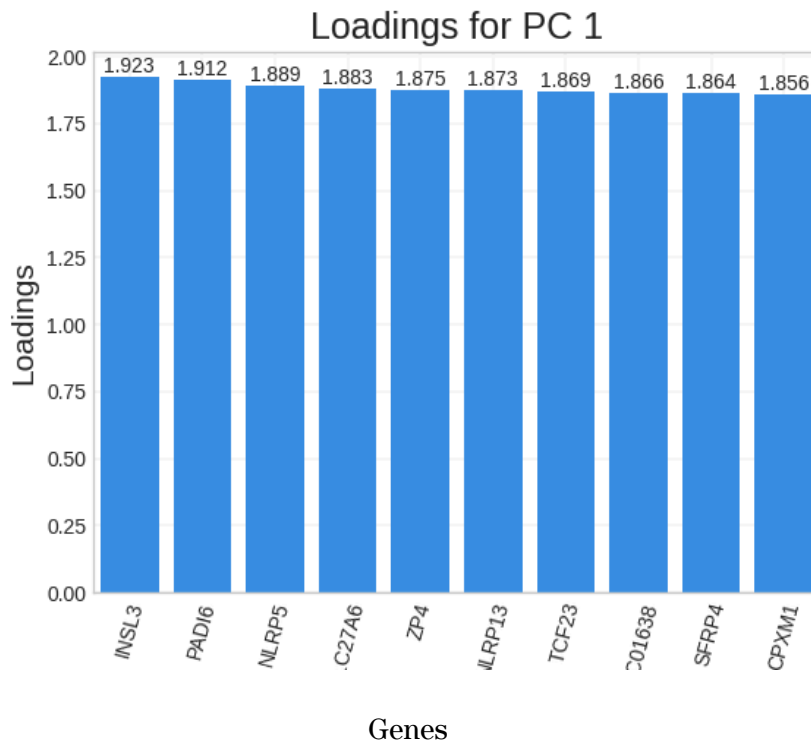
The variances of the first two components are 33% and 21.9%

- Plot the first two principal components (PC1 and PC2) and observe any clustering patterns between tissue groups using PCA Score plot.



The clustering of Lung tissue and ovary tissue can be observed. The tissue from liver and pancreas have clustered together as well

The top ten genes that contribute in the principle component 1 are shown below



2. Differential Gene Expression Analysis (DESeq2):

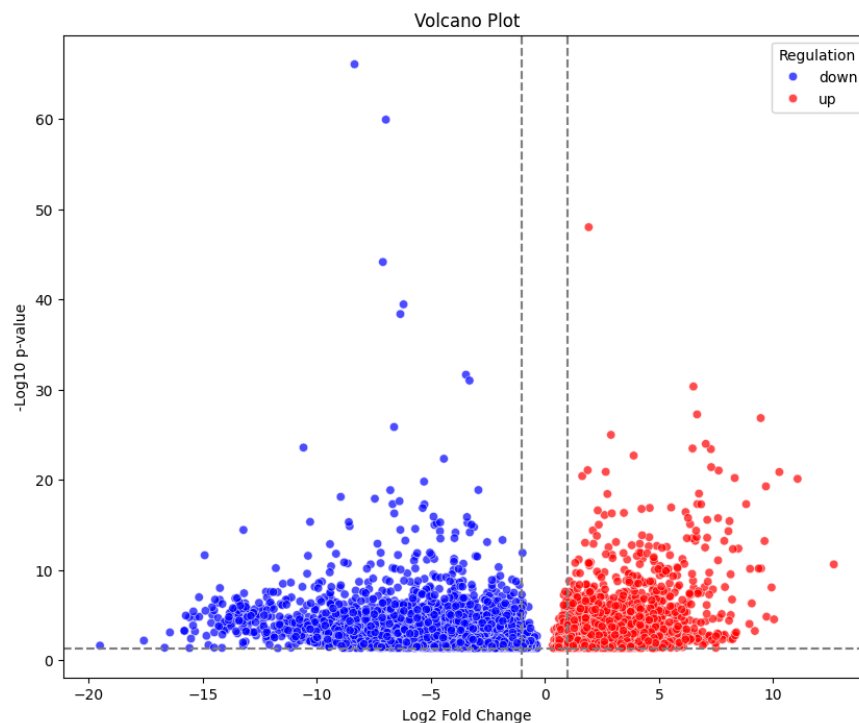
- Using DESeq2, determine which genes are differentially expressed (upregulated or downregulated) in human tissues when comparing ovary tissue against other tissue types.

We edit the metadata file to include only ovary genes and every other genes are considered as others to get the following dataframe

Group	
Gene	
X5EGGE	Others
X5EQMM	Others
X5GICA	Others
X5N9D6	Others
X5P9GB	Others
X5N9ES	Others
X5P9GS	Ovary
X5N9E1	Ovary

The [upregulated](#) and the [downregulated](#) genes are reported in the files.

- Summarize the results by reporting the number of differentially expressed genes and include a volcano plot to illustrate the findings



of genes upregulated = 3071

of genes downregulated = 4746

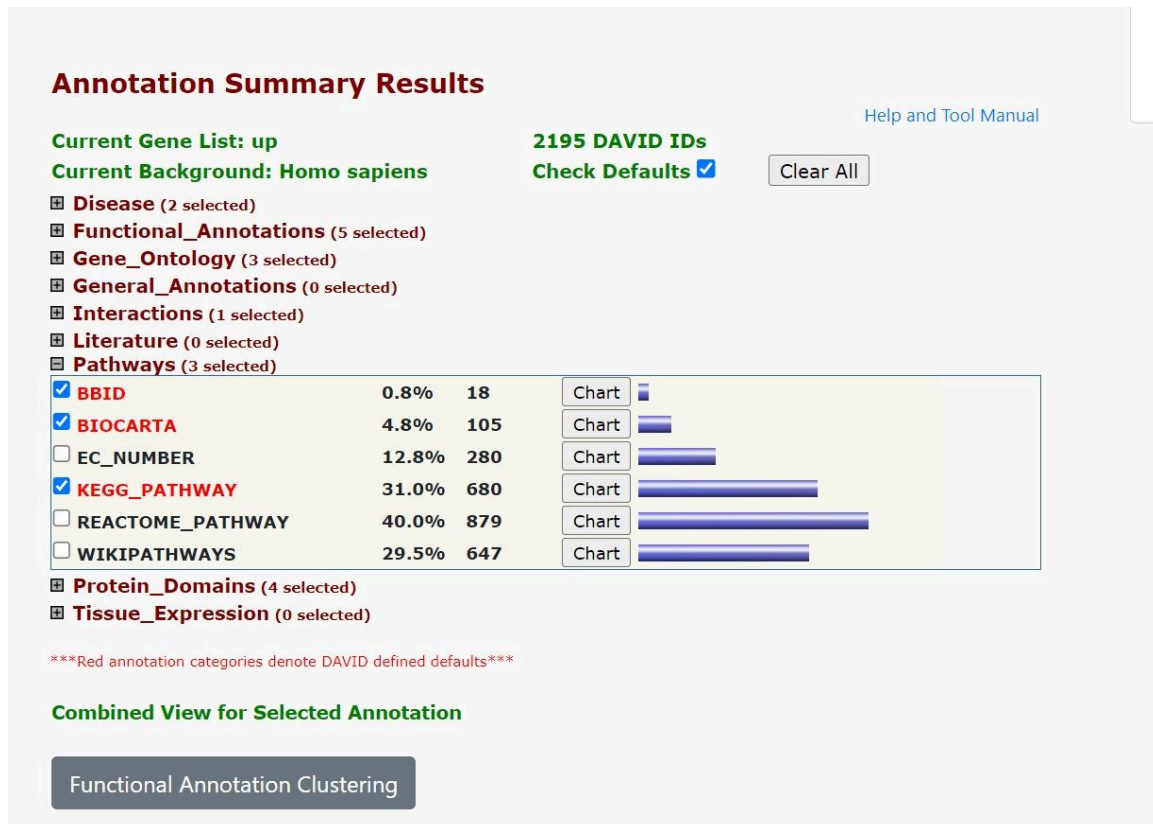
3. Functional Enrichment Analysis (KEGG Pathways):

- Perform a functional enrichment analysis using the KEGG pathway database.

After downloading the upregulated, downregulated and all regulated genes we use the DAVID analyzing tool to perform the functional enrichment analysis.

- Present the significant pathways associated with the upregulated and downregulated genes, highlighting any biological processes of interest.

The pathways associated with upregulated gene are



Functional Annotation Chart

[Help and Manual](#)

Current Gene List: up
Current Background: Homo sapiens
2195 DAVID IDs

Options

Rerun Using Options

Create Sublist

22 chart records

[Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamin
<input type="checkbox"/>	KEGG_PATHWAY	Herpes simplex virus 1 infection	RT		120	5.5	2.2E-29	6.7E-27
<input type="checkbox"/>	KEGG_PATHWAY	Protein digestion and absorption	RT		21	1.0	1.2E-4	1.8E-2
<input type="checkbox"/>	KEGG_PATHWAY	Cytoskeleton in muscle cells	RT		32	1.5	1.7E-3	1.5E-1
<input type="checkbox"/>	KEGG_PATHWAY	Hippo signaling pathway	RT		24	1.1	1.9E-3	1.5E-1
<input type="checkbox"/>	KEGG_PATHWAY	Basal cell carcinoma	RT		13	0.6	2.7E-3	1.6E-1
<input type="checkbox"/>	KEGG_PATHWAY	Hedgehog signaling pathway	RT		12	0.5	3.1E-3	1.6E-1
<input type="checkbox"/>	KEGG_PATHWAY	Ribosome	RT		24	1.1	6.2E-3	2.7E-1
<input type="checkbox"/>	KEGG_PATHWAY	ATP-dependent chromatin remodeling	RT		18	0.8	7.8E-3	2.9E-1
<input type="checkbox"/>	KEGG_PATHWAY	Taste transduction	RT		14	0.6	1.4E-2	4.1E-1
<input type="checkbox"/>	KEGG_PATHWAY	Axon guidance	RT		24	1.1	1.4E-2	4.1E-1
<input type="checkbox"/>	KEGG_PATHWAY	mRNA surveillance pathway	RT		15	0.7	2.5E-2	6.4E-1
<input type="checkbox"/>	KEGG_PATHWAY	Cortisol synthesis and secretion	RT		11	0.5	2.6E-2	6.4E-1
<input type="checkbox"/>	KEGG_PATHWAY	TGF-beta signaling pathway	RT		15	0.7	3.6E-2	8.0E-1
<input type="checkbox"/>	KEGG_PATHWAY	ECM-receptor interaction	RT		13	0.6	3.9E-2	8.0E-1
<input type="checkbox"/>	KEGG_PATHWAY	Coronavirus disease - COVID-19	RT		27	1.2	4.3E-2	8.0E-1
<input type="checkbox"/>	KEGG_PATHWAY	Melanogenesis	RT		14	0.6	4.5E-2	8.0E-1
<input type="checkbox"/>	KEGG_PATHWAY	Morphine addiction	RT		13	0.6	4.5E-2	8.0E-1
<input type="checkbox"/>	KEGG_PATHWAY	Relaxin signaling pathway	RT		16	0.7	7.3E-2	1.0E0
<input type="checkbox"/>	KEGG_PATHWAY	Human papillomavirus infection	RT		34	1.5	7.5E-2	1.0E0
<input type="checkbox"/>	KEGG_PATHWAY	Cushing syndrome	RT		18	0.8	8.6E-2	1.0E0
<input type="checkbox"/>	KEGG_PATHWAY	Ovarian steroidogenesis	RT		8	0.4	9.3E-2	1.0E0
<input type="checkbox"/>	KEGG_PATHWAY	Cell cycle	RT		18	0.8	9.8E-2	1.0E0

The KEGG pathways associated with downregulated genes are

Annotation Summary Results

[Help and Tool Manual](#)

Current Gene List: down

3896 DAVID IDs

Current Background: Homo sapiens

Check Defaults ☒

Clear All

☒ Disease (2 selected)

☒ Functional_Annotations (5 selected)

☒ Gene_Ontology (3 selected)

☒ General_Annotations (0 selected)

☒ Interactions (1 selected)

☒ Literature (0 selected)

☒ Pathways (3 selected)

<input checked="" type="checkbox"/> BBID	2.4%	93	Chart	
<input checked="" type="checkbox"/> BIOCARTA	10.2%	399	Chart	
<input type="checkbox"/> EC_NUMBER	25.6%	997	Chart	
<input checked="" type="checkbox"/> KEGG_PATHWAY	46.4%	1807	Chart	
<input type="checkbox"/> REACTOME_PATHWAY	58.2%	2269	Chart	
<input type="checkbox"/> WIKIPATHWAYS	47.2%	1839	Chart	

☒ Protein_Domains (4 selected)

☒ Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering

Functional Annotation Chart

[Help and Manual](#)

Current Gene List: down

Current Background: Homo sapiens

3896 DAVID IDs

☒ Options

Rerun Using Options

Create Sublist

130 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	KEGG_PATHWAY	Metabolic pathways	RT		441	11.3	3.5E-16	1.2E-13
<input type="checkbox"/>	KEGG_PATHWAY	Complement and coagulation cascades	RT		47	1.2	3.0E-11	5.3E-9
<input type="checkbox"/>	KEGG_PATHWAY	Maturity onset diabetes of the young	RT		21	0.5	9.4E-10	1.1E-7
<input type="checkbox"/>	KEGG_PATHWAY	Retinol metabolism	RT		36	0.9	1.2E-8	1.1E-6
<input type="checkbox"/>	KEGG_PATHWAY	Metabolism of xenobiotics by cytochrome P450	RT		39	1.0	3.1E-8	1.9E-6
<input type="checkbox"/>	KEGG_PATHWAY	Drug metabolism - cytochrome P450	RT		37	0.9	3.1E-8	1.9E-6
<input type="checkbox"/>	KEGG_PATHWAY	Parkinson disease	RT		93	2.4	9.4E-8	4.8E-6
<input type="checkbox"/>	KEGG_PATHWAY	Hematopoietic cell lineage	RT		43	1.1	4.9E-7	2.2E-5
<input type="checkbox"/>	KEGG_PATHWAY	Cytokine-cytokine receptor interaction	RT		97	2.5	7.9E-7	3.1E-5
<input type="checkbox"/>	KEGG_PATHWAY	Fat digestion and absorption	RT		24	0.6	1.6E-6	5.7E-5
<input type="checkbox"/>	KEGG_PATHWAY	Chemical carcinogenesis - DNA adducts	RT		33	0.8	2.3E-6	7.5E-5
<input type="checkbox"/>	KEGG_PATHWAY	Prion disease	RT		90	2.3	2.7E-6	7.9E-5
<input type="checkbox"/>	KEGG_PATHWAY	Viral protein interaction with cytokine and cytokine receptor	RT		41	1.1	5.4E-6	1.4E-4
<input type="checkbox"/>	KEGG_PATHWAY	Bile secretion	RT		38	1.0	5.7E-6	1.4E-4
<input type="checkbox"/>	KEGG_PATHWAY	Alzheimer disease	RT		117	3.0	5.9E-6	1.4E-4
<input type="checkbox"/>	KEGG_PATHWAY	Diabetic cardiomyopathy	RT		69	1.8	1.1E-5	2.4E-4
<input type="checkbox"/>	KEGG_PATHWAY	Chemical carcinogenesis - reactive oxygen species	RT		74	1.9	1.5E-5	3.1E-4
<input type="checkbox"/>	KEGG_PATHWAY	Cholesterol metabolism	RT		25	0.6	1.7E-5	3.4E-4
<input type="checkbox"/>	KEGG_PATHWAY	Chemokine signaling pathway	RT		65	1.7	2.0E-5	3.7E-4
<input type="checkbox"/>	KEGG_PATHWAY	Non-alcoholic fatty liver disease	RT		55	1.4	2.8E-5	4.5E-4
<input type="checkbox"/>	KEGG_PATHWAY	Cell adhesion molecules	RT		55	1.4	2.8E-5	4.5E-4
<input type="checkbox"/>	KEGG_PATHWAY	Pancreatic secretion	RT		41	1.1	2.8E-5	4.5E-4
<input type="checkbox"/>	KEGG_PATHWAY	Lipid and atherosclerosis	RT		69	1.8	7.1E-5	1.1E-3
<input type="checkbox"/>	KEGG_PATHWAY	Tyrosine metabolism	RT		19	0.5	7.2E-5	1.1E-3
<input type="checkbox"/>	KEGG_PATHWAY	Proteasome	RT		22	0.6	1.0E-4	1.4E-3
<input type="checkbox"/>	KEGG_PATHWAY	Pathogenic Escherichia coli infection	RT		65	1.7	1.1E-4	1.5E-3
<input type="checkbox"/>	KEGG_PATHWAY	B cell receptor signalling pathway	RT		35	0.9	1.4E-4	1.8E-3

For all genes that are regulated, we find the following KEGG pathways

Annotation Summary Results

Current Gene List: normal

Current Background: Homo sapiens

6091 DAVID IDs

Check Defaults ☒

[Help and Tool Manual](#)

Clear All

- ☒ **Disease** (2 selected)
- ☒ **Functional_Annotations** (5 selected)
- ☒ **Gene_Ontology** (3 selected)
- ☒ **General_Annotations** (0 selected)
- ☒ **Interactions** (1 selected)
- ☒ **Literature** (0 selected)
- ☒ **Pathways** (3 selected)

<input checked="" type="checkbox"/> BBID	1.8%	111	Chart	
<input checked="" type="checkbox"/> BIOCARTA	8.3%	504	Chart	
<input type="checkbox"/> EC_NUMBER	21.0%	1277	Chart	
<input checked="" type="checkbox"/> KEGG_PATHWAY	40.8%	2487	Chart	
<input type="checkbox"/> REACTOME_PATHWAY	51.7%	3148	Chart	
<input type="checkbox"/> WIKIPATHWAYS	40.8%	2486	Chart	

- ☒ **Protein_Domains** (4 selected)
- ☒ **Tissue_Expression** (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering

Functional Annotation Chart

Current Gene List: normal

Current Background: Homo sapiens

6091 DAVID IDs

Options

Rerun Using Options

Create Sublist

95 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamin
<input type="checkbox"/>	KEGG_PATHWAY	Complement and coagulation cascades	RT		49	0.8	1.5E-7	5.4E-5
<input type="checkbox"/>	KEGG_PATHWAY	Maturity onset diabetes of the young	RT		21	0.3	3.3E-7	5.8E-5
<input type="checkbox"/>	KEGG_PATHWAY	Protein digestion and absorption	RT		53	0.9	2.5E-6	2.9E-4
<input type="checkbox"/>	KEGG_PATHWAY	Metabolism of xenobiotics by cytochrome P450	RT		42	0.7	6.5E-6	4.7E-4
<input type="checkbox"/>	KEGG_PATHWAY	Cell adhesion molecules	RT		71	1.2	6.6E-6	4.7E-4
<input type="checkbox"/>	KEGG_PATHWAY	Drug metabolism - cytochrome P450	RT		39	0.6	1.3E-5	7.6E-4
<input type="checkbox"/>	KEGG_PATHWAY	Metabolic pathways	RT		506	8.3	2.8E-5	1.4E-3
<input type="checkbox"/>	KEGG_PATHWAY	Fat digestion and absorption	RT		26	0.4	3.5E-5	1.5E-3
<input type="checkbox"/>	KEGG_PATHWAY	Retinol metabolism	RT		36	0.6	3.8E-5	1.5E-3
<input type="checkbox"/>	KEGG_PATHWAY	Parkinson disease	RT		106	1.7	8.0E-5	2.9E-3
<input type="checkbox"/>	KEGG_PATHWAY	Bile secretion	RT		43	0.7	1.3E-4	4.2E-3
<input type="checkbox"/>	KEGG_PATHWAY	Chemical carcinogenesis - DNA adducts	RT		35	0.6	3.0E-4	9.0E-3
<input type="checkbox"/>	KEGG_PATHWAY	Prion disease	RT		105	1.7	4.1E-4	1.1E-2
<input type="checkbox"/>	KEGG_PATHWAY	Pancreatic secretion	RT		47	0.8	5.1E-4	1.3E-2
<input type="checkbox"/>	KEGG_PATHWAY	Hematopoietic cell lineage	RT		44	0.7	7.4E-4	1.8E-2
<input type="checkbox"/>	KEGG_PATHWAY	Cholesterol metabolism	RT		26	0.4	1.2E-3	2.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	Chemokine signaling pathway	RT		75	1.2	1.2E-3	2.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	Alzheimer disease	RT		138	2.3	1.4E-3	2.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	PPAR signaling pathway	RT		35	0.6	1.4E-3	2.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	Tight junction	RT		67	1.1	1.5E-3	2.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	Tyrosine metabolism	RT		20	0.3	1.5E-3	2.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	Arginine and proline metabolism	RT		25	0.4	2.2E-3	3.5E-2
<input type="checkbox"/>	KEGG_PATHWAY	Pathogenic Escherichia coli infection	RT		77	1.3	2.3E-3	3.5E-2
<input type="checkbox"/>	KEGG_PATHWAY	Protein processing in endoplasmic reticulum	RT		66	1.1	2.5E-3	3.8E-2
<input type="checkbox"/>	KEGG_PATHWAY	Diabetic cardiomyopathy	RT		77	1.3	3.0E-3	4.4E-2
<input type="checkbox"/>	KEGG_PATHWAY	B cell receptor signaling pathway	RT		39	0.6	3.4E-3	4.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	Proteasome	RT		23	0.4	3.5E-3	4.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	Cytokine-cytokine receptor interaction	RT		106	1.7	3.9E-3	4.9E-2
<input type="checkbox"/>	KEGG_PATHWAY	Chemical carcinogenesis - reactive oxygen species	RT		83	1.4	4.2E-3	5.2E-2
<input type="checkbox"/>	KEGG_PATHWAY	Non-alcoholic fatty liver disease	RT		60	1.0	6.0E-3	7.2E-2
<input type="checkbox"/>	KEGG_PATHWAY	Steroid hormone biosynthesis	RT		28	0.5	6.5E-3	7.4E-2

The significant pathways with the lowest p-value for

1) Downregulated genes

Metabolic pathways

Complement and coagulation cascades

Maturity onset diabetes of the young

Retinol Metabolism

2) Upregulated genes

Herpes simplex virus infection

Protein digestion and absorption

Cytoskeleton in muscle cells

Hippo signaling pathway

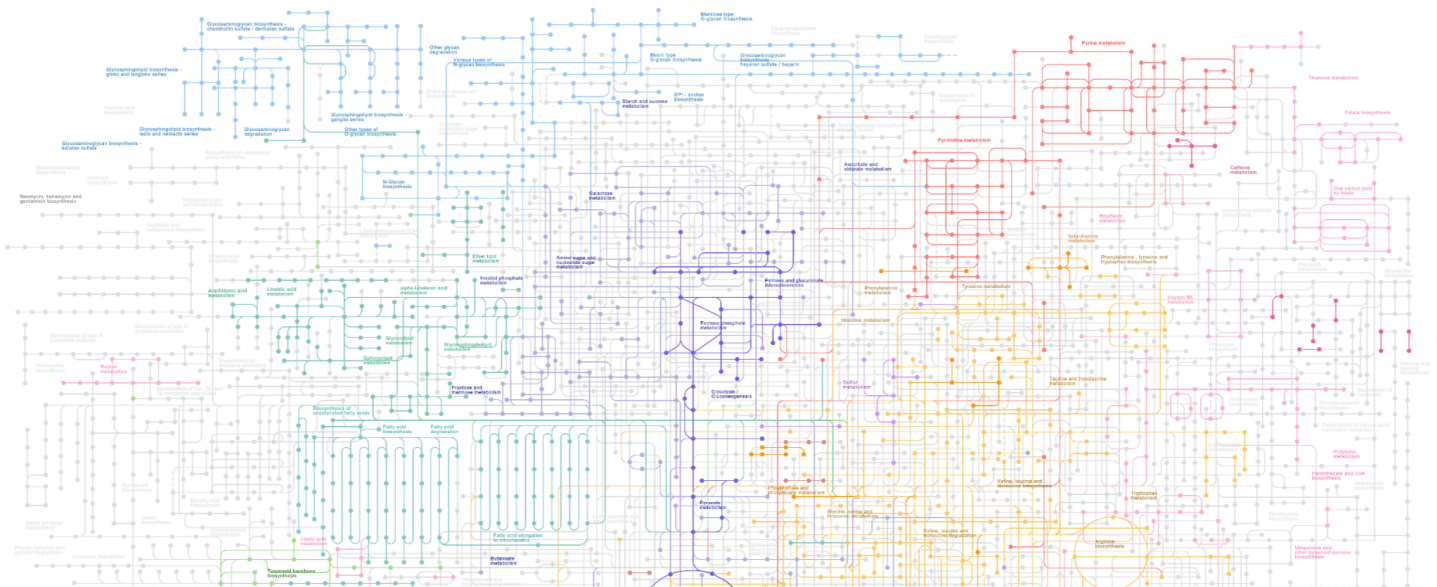
3) For all genes

Complement and coagulation cascades

Maturity onset diabetes of the young

Protein digestion and absorption

The metabolic pathway



Herpes simplex virus infection



DAVID Gene Name
2'-5'-oligoadenylate synthetase 1(OAS1)

DAVID Gene Name

2'-5'-oligoadenylate synthetase 1 (OAS1)