BioinformaticsHW_IsabelMetzger

ISABEL METZGER BIOINFORMATICS FALL 2017 LOADING EDGER

```
library(limma)
```

```
## Warning: package 'limma' was built under R version 3.4.2
```

```
library(edgeR)
library(car)
```

Read in count table and experimental design:

```
ovarian_df <- read.delim("GSE52695_Ovarian_RefSeq_RPKM_Values.txt",
                         row.names=1, header=T)
```

```
head(ovarian_df)
```

```
##          Library_1 Library_10 Library_11 Library_12  Library_2 Library_3
## A1BG     0.2621930  0.9163250 0.90431500 1.03555000 0.30342000 0.3397060
## A1BG-AS1 0.9704190  0.8373180 1.38764000 1.51185000 0.66485800 1.0530700
## A1CF     0.0103266  0.0212174 0.03549150 0.00425699 0.01229490 0.0111711
## A2LD1    0.5142540  1.5253800 0.92915700 1.10681000 0.41714800 0.4988250
## A2M      0.0000000  0.0283904 0.00647592 0.00854420 0.00822571 0.0000000
## A2ML1    0.0000000  0.0127313 0.01161620 0.00766310 0.00737746 0.0000000
##           Library_4  Library_5  Library_6 Library_7  Library_8  Library_9
## A1BG     0.44528100 0.51423700 0.30590300 0.6488740 0.65693400 0.81654700
## A1BG-AS1 0.68563300 0.64529800 0.99422200 1.0894200 0.99634600 0.86571700
## A1CF     0.00902164 0.01051260 0.00925534 0.0214581 0.01211950 0.02153980
## A2LD1    0.65327200 0.40458300 0.50209300 0.5161380 0.42464100 1.39639000
## A2M      0.37421800 0.00527496 0.00928820 0.0000000 0.00608129 0.00000000
## A2ML1    0.00000000 0.00000000 0.00000000 0.0000000 0.00000000 0.00646238
```

the dataframe contains 22446 rows (genes) and a total of 12 cols (12 libraries)

```
dim(ovarian_df)
```

```
## [1] 22446    12
```

Reordering column names

```
reorder.colnames <- paste("Library", 1:12, sep="_")
```

```
reorder.colnames
```

```
##  [1] "Library_1"  "Library_2"  "Library_3"  "Library_4"  "Library_5"
##  [6] "Library_6"  "Library_7"  "Library_8"  "Library_9"  "Library_10"
## [11] "Library_11" "Library_12"
```

```
ovarian_reordered_df <- ovarian_df[,reorder.colnames]
```

```
head(ovarian_reordered_df)
```

```
##          Library_1  Library_2 Library_3  Library_4  Library_5  Library_6
## A1BG     0.2621930 0.30342000 0.3397060 0.44528100 0.51423700 0.30590300
## A1BG-AS1 0.9704190 0.66485800 1.0530700 0.68563300 0.64529800 0.99422200
```

```
## A1CF     0.0103266 0.01229490 0.0111711 0.00902164 0.01051260 0.00925534
## A2LD1    0.5142540 0.41714800 0.4988250 0.65327200 0.40458300 0.50209300
## A2M      0.0000000 0.00822571 0.0000000 0.37421800 0.00527496 0.00928820
## A2ML1    0.0000000 0.00737746 0.0000000 0.00000000 0.00000000 0.00000000
##          Library_7  Library_8  Library_9 Library_10 Library_11 Library_12
## A1BG     0.6488740 0.65693400 0.81654700  0.9163250 0.90431500 1.03555000
## A1BG-AS1 1.0894200 0.99634600 0.86571700  0.8373180 1.38764000 1.51185000
## A1CF     0.0214581 0.01211950 0.02153980  0.0212174 0.03549150 0.00425699
## A2LD1    0.5161380 0.42464100 1.39639000  1.5253800 0.92915700 1.10681000
## A2M      0.0000000 0.00608129 0.00000000  0.0283904 0.00647592 0.00854420
## A2ML1    0.0000000 0.00000000 0.00646238  0.0127313 0.01161620 0.00766310
```

Creating new colnames

```
newcolnames <- paste("GSM12742", 53:64, sep="") #2 or more vectors pasted element for element.
```

```
newcolnames
```

```
##  [1] "GSM1274253" "GSM1274254" "GSM1274255" "GSM1274256" "GSM1274257"
##  [6] "GSM1274258" "GSM1274259" "GSM1274260" "GSM1274261" "GSM1274262"
## [11] "GSM1274263" "GSM1274264"
```

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.4.2
```

```
new_ovarian_df <- setnames(ovarian_reordered_df,
                           old=reorder.colnames, new=newcolnames)
```

```
head(new_ovarian_df)
```

```
##          GSM1274253 GSM1274254 GSM1274255 GSM1274256 GSM1274257 GSM1274258
## A1BG      0.2621930 0.30342000  0.3397060 0.44528100 0.51423700 0.30590300
## A1BG-AS1  0.9704190 0.66485800  1.0530700 0.68563300 0.64529800 0.99422200
## A1CF      0.0103266 0.01229490  0.0111711 0.00902164 0.01051260 0.00925534
## A2LD1     0.5142540 0.41714800  0.4988250 0.65327200 0.40458300 0.50209300
## A2M       0.0000000 0.00822571  0.0000000 0.37421800 0.00527496 0.00928820
## A2ML1     0.0000000 0.00737746  0.0000000 0.00000000 0.00000000 0.00000000
##          GSM1274259 GSM1274260 GSM1274261 GSM1274262 GSM1274263 GSM1274264
## A1BG      0.6488740 0.65693400 0.81654700  0.9163250 0.90431500 1.03555000
## A1BG-AS1  1.0894200 0.99634600 0.86571700  0.8373180 1.38764000 1.51185000
## A1CF      0.0214581 0.01211950 0.02153980  0.0212174 0.03549150 0.00425699
## A2LD1     0.5161380 0.42464100 1.39639000  1.5253800 0.92915700 1.10681000
## A2M       0.0000000 0.00608129 0.00000000  0.0283904 0.00647592 0.00854420
## A2ML1     0.0000000 0.00000000 0.00646238  0.0127313 0.01161620 0.00766310
```

Genes are now row names and what is left is the counts/ expression levels. See below for the renamed columns.

```
length(reorder.colnames)
```

```
## [1] 12
```

```
length(newcolnames)
```

```
## [1] 12
```

```
treatment <- c("Normoxia_6hr (total RNA)_1", "Normoxia_6hr (total RNA)_2",
               "Hypoxia_6hr (total RNA)_3","Hypoxia_6hr (total RNA)_4","Normoxia_48hr (total RNA)_5","
```

```
              "Hypoxia_48hr (total RNA)_8","Normoxia_6days (total RNA)_9","Normoxia_6days (total RNA)_1
length(treatment)
```

```
## [1] 12
```

```
keydf <- data.frame(reorder.colnames, newcolnames, treatment)
```

```
keydf$treatmentsimplified <- c("Normoxia", "Normoxia", "Hypoxia", "Hypoxia","Normoxia", "Normoxia", "Hyp
```

```
keydf
```

```
##    reorder.colnames newcolnames                    treatment
## 1         Library_1  GSM1274253    Normoxia_6hr (total RNA)_1
## 2         Library_2  GSM1274254    Normoxia_6hr (total RNA)_2
## 3         Library_3  GSM1274255     Hypoxia_6hr (total RNA)_3
## 4         Library_4  GSM1274256     Hypoxia_6hr (total RNA)_4
## 5         Library_5  GSM1274257   Normoxia_48hr (total RNA)_5
## 6         Library_6  GSM1274258   Normoxia_48hr (total RNA)_6
## 7         Library_7  GSM1274259    Hypoxia_48hr (total RNA)_7
## 8         Library_8  GSM1274260    Hypoxia_48hr (total RNA)_8
## 9         Library_9  GSM1274261  Normoxia_6days (total RNA)_9
## 10       Library_10  GSM1274262 Normoxia_6days (total RNA)_10
## 11       Library_11  GSM1274263  Hypoxia_6days (total RNA)_11
## 12       Library_12  GSM1274264  Hypoxia_6days (total RNA)_12
##    treatmentsimplified
## 1             Normoxia
## 2             Normoxia
## 3              Hypoxia
## 4              Hypoxia
## 5             Normoxia
## 6             Normoxia
## 7              Hypoxia
## 8              Hypoxia
## 9             Normoxia
## 10            Normoxia
## 11             Hypoxia
## 12             Hypoxia
```

```
group <- keydf$treatmentsimplified
```

```
group
```

```
##  [1] "Normoxia" "Normoxia" "Hypoxia"  "Hypoxia"  "Normoxia" "Normoxia"
##  [7] "Hypoxia"  "Hypoxia"  "Normoxia" "Normoxia" "Hypoxia"  "Hypoxia"
```

```
head(new_ovarian_df)
```

```
##          GSM1274253 GSM1274254 GSM1274255 GSM1274256 GSM1274257 GSM1274258
## A1BG      0.2621930 0.30342000  0.3397060 0.44528100 0.51423700 0.30590300
## A1BG-AS1  0.9704190 0.66485800  1.0530700 0.68563300 0.64529800 0.99422200
## A1CF      0.0103266 0.01229490  0.0111711 0.00902164 0.01051260 0.00925534
## A2LD1     0.5142540 0.41714800  0.4988250 0.65327200 0.40458300 0.50209300
## A2M       0.0000000 0.00822571  0.0000000 0.37421800 0.00527496 0.00928820
## A2ML1     0.0000000 0.00737746  0.0000000 0.00000000 0.00000000 0.00000000
##          GSM1274259 GSM1274260 GSM1274261 GSM1274262 GSM1274263 GSM1274264
## A1BG      0.6488740 0.65693400 0.81654700  0.9163250 0.90431500 1.03555000
## A1BG-AS1  1.0894200 0.99634600 0.86571700  0.8373180 1.38764000 1.51185000
```

```
## A1CF      0.0214581 0.01211950 0.02153980  0.0212174 0.03549150 0.00425699
## A2LD1     0.5161380 0.42464100 1.39639000  1.5253800 0.92915700 1.10681000
## A2M       0.0000000 0.00608129 0.00000000  0.0283904 0.00647592 0.00854420
## A2ML1     0.0000000 0.00000000 0.00646238  0.0127313 0.01161620 0.00766310
```

```
#Create DGEList object with groups for the treatments N and H

cds <- DGEList(new_ovarian_df, group = group )
dim(cds)
```

```
## [1] 22446    12
```

```
#Filter out genes with low counts, keeping those rows where the count
#per million (cpm) is at least 1 in at least three samples:

keep <- rowSums(cpm(cds)>1) >=3
cds <- cds[keep,]
dim(cds)                # How many genes are left?
```

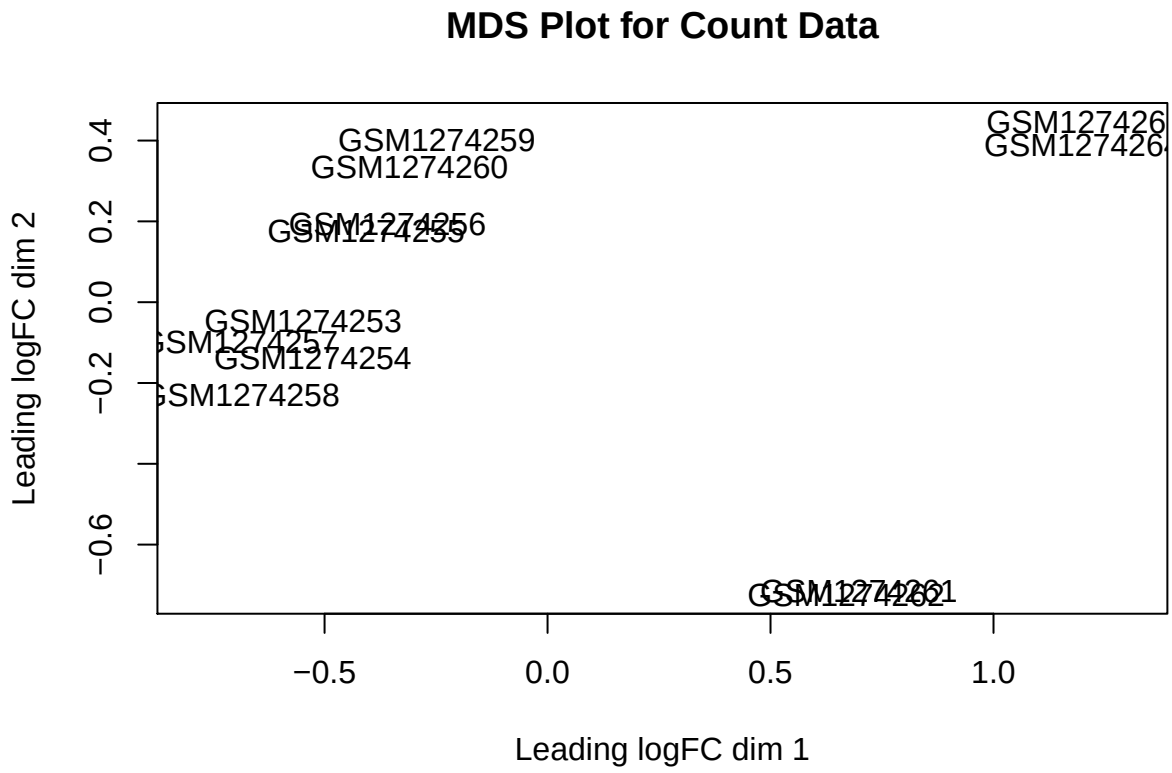
```
## [1] 12162    12
```

```
cds <- calcNormFactors(cds) #normalize number of reads per sample
```

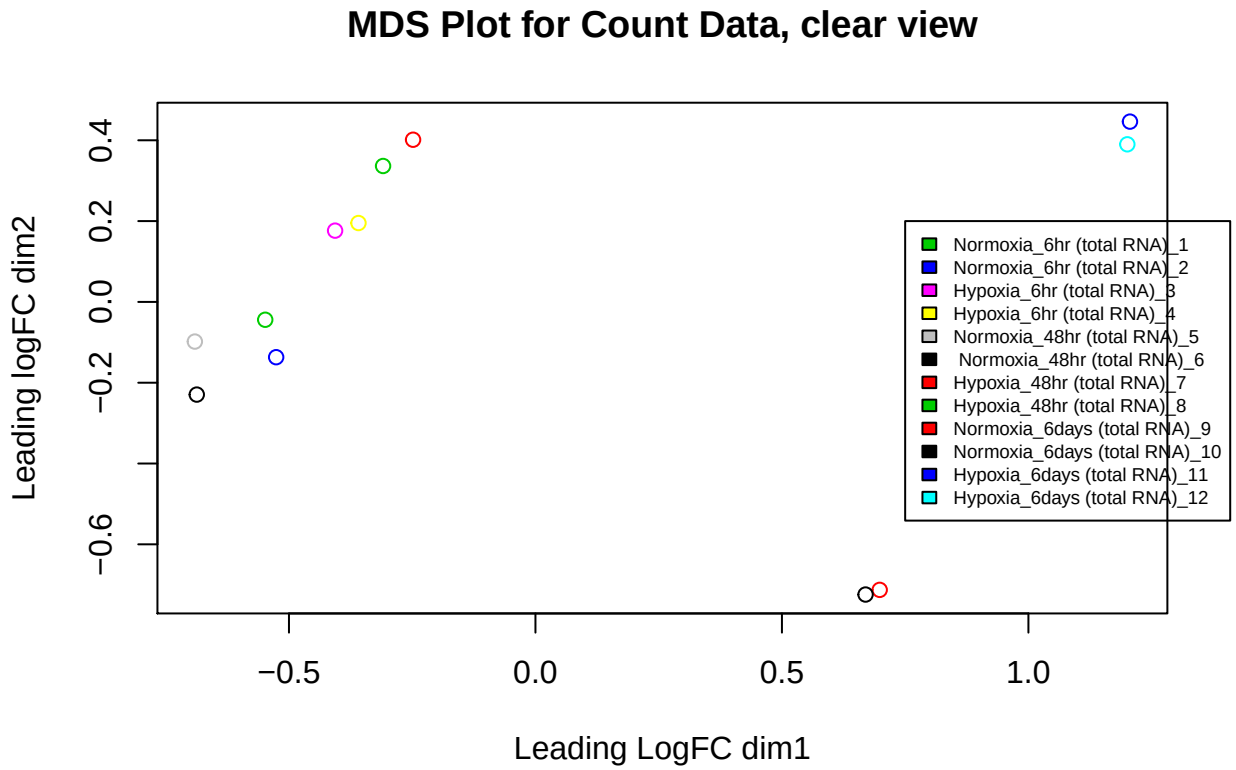
```
?plotMDS
```

Usage ## Default S3 method: plotMDS(x, top = 500, labels = NULL, pch = NULL, cex = 1, dim.plot = c(1,2), ndim = max(dim.plot), gene.selection = "pairwise", xlab = NULL, ylab = NULL, plot = TRUE, ...) ## S3 method for class 'MDS' plotMDS(x, labels = NULL, pch = NULL, cex = 1, dim.plot = NULL, xlab = NULL, ylab = NULL, ...)

```
               #plot the data (MDS is like PCA)
mds <- plotMDS(cds, main = "MDS Plot for Count Data", labels = colnames(cds$counts))
```



```
plot(mds, main="MDS Plot for Count Data, clear view", xlab="Leading LogFC dim1", ylab="Leading logFC dim
par(xpd=TRUE)
legend(0.75,0.2,fill=keydf$treatment,legend=keydf$treatment,cex=0.6)
```




```
#           title("Tx Status")
# library(car)
```

```
cds
```

```
## An object of class "DGEList"
## $counts
##          GSM1274253 GSM1274254 GSM1274255 GSM1274256 GSM1274257 GSM1274258
## A1BG       0.262193   0.303420   0.339706   0.445281   0.514237   0.305903
## A1BG-AS1   0.970419   0.664858   1.053070   0.685633   0.645298   0.994222
## A2LD1      0.514254   0.417148   0.498825   0.653272   0.404583   0.502093
## AAAS       8.223290   7.361950   9.427670   7.773020  10.328000  10.390200
## AACS       5.169470   4.668560   5.389730   4.583680   4.632230   4.953550
##          GSM1274259 GSM1274260 GSM1274261 GSM1274262 GSM1274263 GSM1274264
## A1BG       0.648874   0.656934   0.816547   0.916325   0.904315    1.03555
## A1BG-AS1   1.089420   0.996346   0.865717   0.837318   1.387640    1.51185
## A2LD1      0.516138   0.424641   1.396390   1.525380   0.929157    1.10681
## AAAS      10.274000   9.328310   7.662060   8.270740   6.487310    6.94087
## AACS       5.533260   5.850510   3.634350   3.553840   5.642410    5.87783
## 12157 more rows ...
##
## $samples
##                group lib.size norm.factors
## GSM1274253 Normoxia 772461.6    0.8880870
## GSM1274254 Normoxia 603231.3    1.0442287
## GSM1274255  Hypoxia 618097.6    1.1320669
```

```
## GSM1274256  Hypoxia 615240.6    1.0903608
## GSM1274257 Normoxia 788218.0    0.8841033
## 7 more rows ...
```

```
               #estimate variance, compute diff. exp
cds <- estimateCommonDisp( cds )
cds <- estimateTagwiseDisp( cds , prior.df = 10 )   #shrinks variance toward common disp.
de.tgw <- exactTest( cds , pair = c( "Normoxia" , "Hypoxia" ) ) #this is the DE test
options( digits = 3 ) # print only 3 digits
topTags( de.tgw , n = 20 , sort.by = "p.value" )      # print the top 20 DE genes, by p-value
```

```
## Comparison of groups:  Hypoxia-Normoxia
##         logFC logCPM   PValue      FDR
## PFKFB4   3.37   4.19 3.73e-23 4.54e-19
## BNIP3    2.62   6.88 8.05e-17 4.90e-13
## SLC2A1   2.28   4.30 7.96e-16 3.23e-12
## HK2      1.83   5.06 1.08e-13 3.29e-10
## ADM      2.50   4.17 1.85e-13 4.49e-10
## ANGPTL4  3.11   3.75 2.43e-13 4.93e-10
## SLC16A3  1.90   4.86 4.07e-13 7.07e-10
## ANKZF1   1.77   4.47 2.99e-12 4.54e-09
## FAM162A  1.85   6.83 4.35e-12 5.88e-09
## AK4      1.77   4.42 5.46e-12 6.63e-09
## PFKP     1.87   5.17 1.50e-11 1.66e-08
## PGK1     1.53   7.94 1.67e-11 1.69e-08
## INSIG2   2.44   3.99 3.12e-11 2.91e-08
## P4HA2    1.80   5.13 3.39e-11 2.95e-08
## EGLN1    1.60   4.58 6.86e-11 5.56e-08
## PLOD2    2.32   4.99 7.53e-11 5.73e-08
## TMEM45A  2.25   4.19 4.06e-10 2.91e-07
## ENO2     2.24   5.72 8.25e-10 5.47e-07
## PPFIA4   3.33   3.14 8.55e-10 5.47e-07
## PFKFB3   1.49   4.44 2.25e-09 1.37e-06
```

```
resultsTbl.tgw <- topTags( de.tgw , n = nrow( de.tgw$table ) )$table
de.genes.tgw <- rownames( resultsTbl.tgw )[ resultsTbl.tgw$FDR <= 0.05 ]
length( de.genes.tgw )      # How many genes with significant DE?
```

```
## [1] 105
```

```
summary( decideTestsDGE( de.tgw , p.value = 0.05 ) )        # show Up and Down regulated genes
```

```
##    Normoxia+Hypoxia
## -1                1
## 0             12057
## 1               104
```

```
write.table(resultsTbl.tgw, file = "LiData_egeR_.csv" , sep = "," , row.names = TRUE )
```

```
# xgenes <- cat(noquote(row.names(resultsTbl.tgw)), sep="\n")
```

Took list of top DE genes and used DAVID to find Gene Ontology and KEGG enriched functional groups. Went to website: https://david.ncifcrf.gov/summary.jsp Uploaded gene list, selected Homo sapiens:

I picked the top 225 genes, selected Homo sapiens, which narrowed the Gene List to 213 genes.

Here is an image of the process: Explored various ways to import the URL tab delim file.

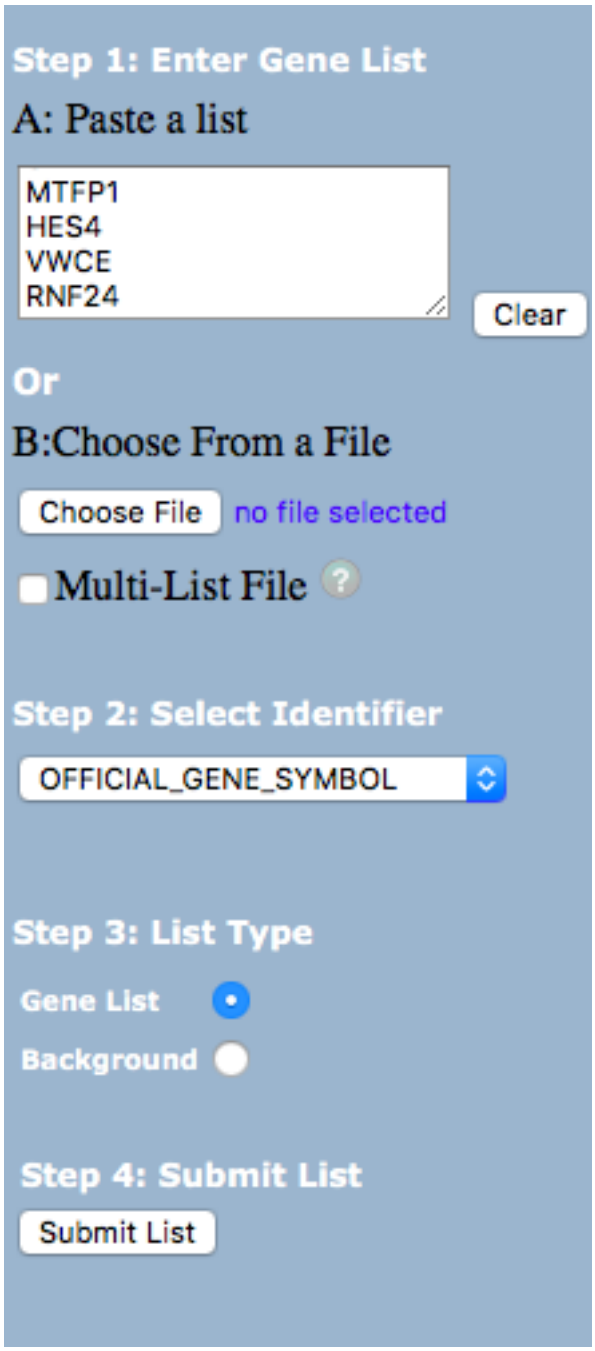


Figure 1: Pasting gene list.

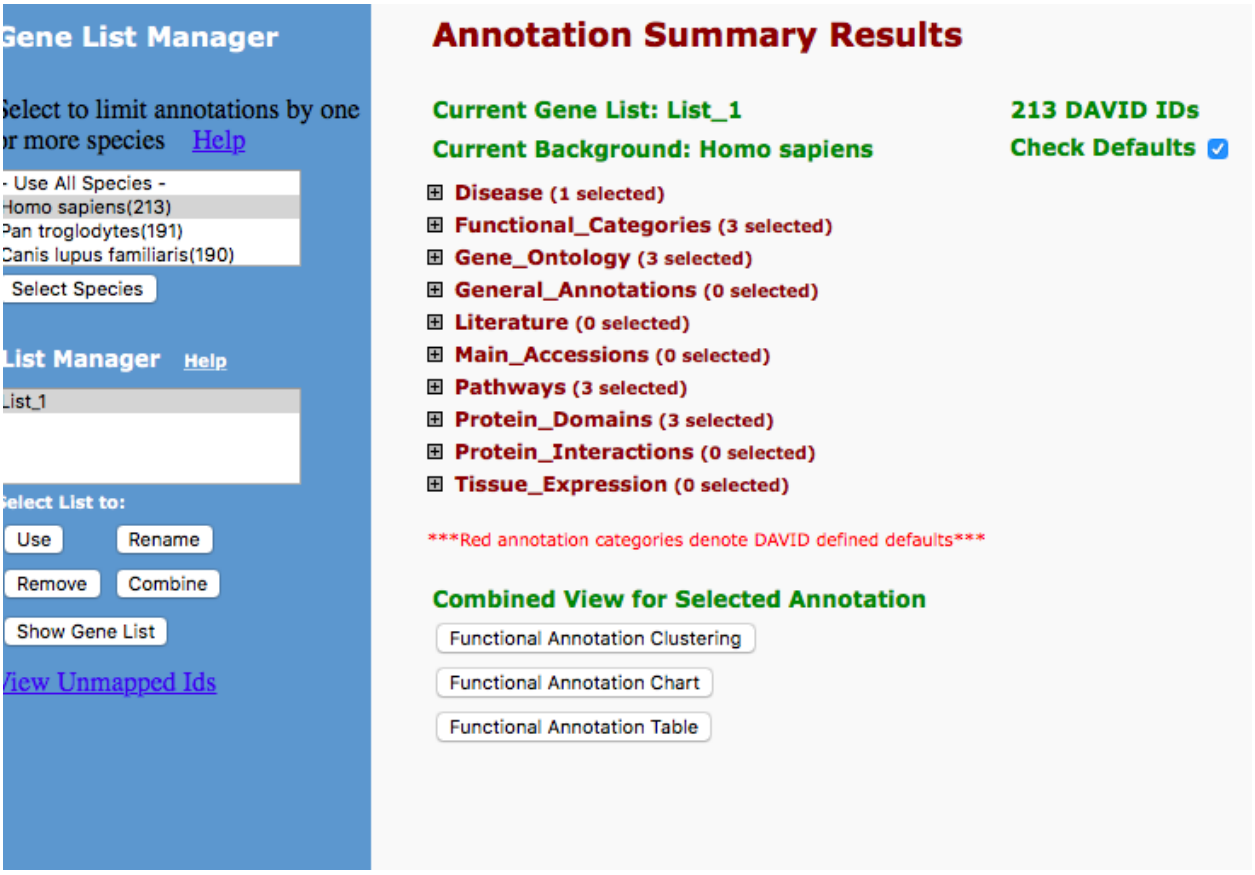


Figure 2: Selecting spcies.

```
# ![Selecting spcies.](selectingspecies.png)
```

Current Gene List: List_1 Current Background: Homo sapiens 213 DAVID IDs ——— Importing the functional_annotation_clustering for *Homo sapiens*

```
library(readr)
library(RCurl)
```

```
## Loading required package: bitops
```

```
  URL <-"https://david.ncifcrf.gov/data/download/t2t_0C32396F8F541508972082264.txt"

# Final_Clustering <- read.table(URL, header=FALSE, sep="\t")
```

```
functional.annotaton.clustering.dat <- fread(URL, header=T)
```

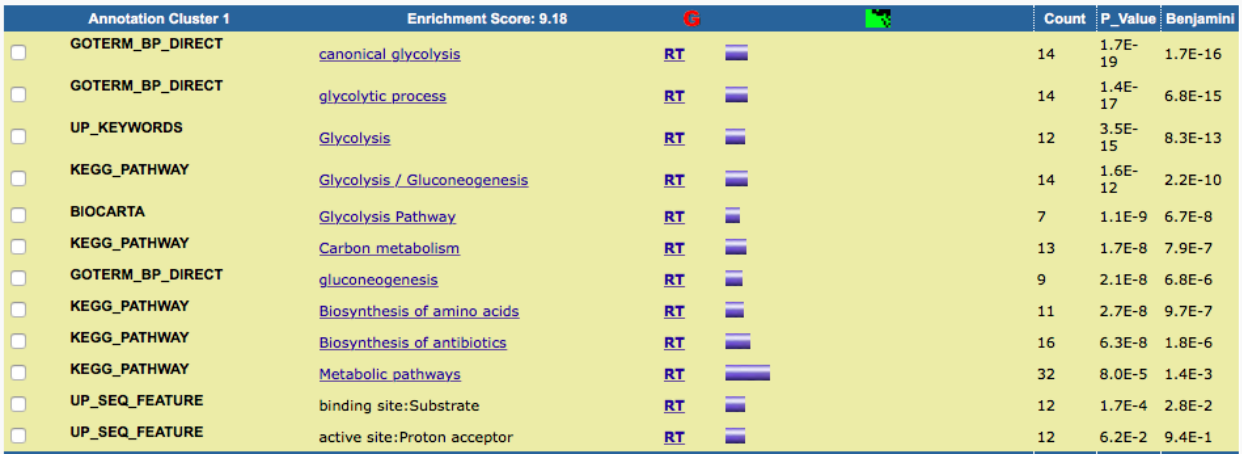
```
## Warning in fread(URL, header = T): Starting data input on line 2 and
## discarding line 1 because it has too few or too many items to be column
## names or data: Annotation Cluster 1 Enrichment Score: 9.178650479827342
```

```
## Warning in fread(URL, header = T): Stopped reading at empty line 15 but
## text exists afterwards (discarded): Annotation Cluster 2 Enrichment Score:
## 4.539764866201056
```

```
head(functional.annotaton.clustering.dat)
```

```
##            Category                                     Term Count    %
## 1: GOTERM_BP_DIRECT        GO:0061621~canonical glycolysis    14 6.57
## 2: GOTERM_BP_DIRECT          GO:0006096~glycolytic process    14 6.57
## 3:      UP_KEYWORDS                              Glycolysis    12 5.63
## 4:     KEGG_PATHWAY  hsa00010:Glycolysis / Gluconeogenesis    14 6.57
## 5:         BIOCARTA h_glycolysisPathway:Glycolysis Pathway     7 3.29
## 6:     KEGG_PATHWAY             hsa01200:Carbon metabolism    13 6.10
##      PValue
## 1: 1.68e-19
## 2: 1.38e-17
## 3: 3.48e-15
## 4: 1.57e-12
## 5: 1.07e-09
## 6: 1.68e-08
##                                                                                Genes
## 1: ALDOA, PFKL, PFKFB4, PFKFB3, ALDOC, HK2, PGAM1, PFKP, HK1, TPI1, ENO2, PGK1, GAPDH, ENO1
## 2:  ALDOA, LDHA, PFKL, ALDOC, HK2, PGAM1, HK1, TPI1, PGM1, ENO2, PGK1, GAPDH, EDARADD, ENO1
## 3:                 ALDOA, TPI1, PFKL, ALDOC, ENO2, PGAM1, HK2, PFKP, HK1, PGK1, GAPDH, ENO1
## 4:     ALDOA, LDHA, PFKL, ALDOC, HK2, PGAM1, PFKP, HK1, TPI1, PGM1, ENO2, PGK1, GAPDH, ENO1
## 5:                                                TPI1, PFKL, PGAM1, HK1, PGK1, GAPDH, ENO1
## 6:           ALDOA, TPI1, PFKL, ALDOC, ENO2, PGAM1, HK2, PFKP, HK1, PGK1, GPT2, GAPDH, ENO1
##    List Total Pop Hits Pop Total Fold Enrichment Bonferroni Benjamini
## 1:        183       26     16792           49.41   1.65e-16  1.65e-16
## 2:        183       34     16792           37.78   1.36e-14  6.80e-15
## 3:        199       31     20581           40.03   8.29e-13  8.29e-13
## 4:         90       67      6910           16.04   2.22e-10  2.22e-10
## 5:         25       10      1625           45.50   6.66e-08  6.66e-08
## 6:         90      113      6910            8.83   2.36e-06  7.88e-07
##         FDR
## 1: 2.65e-16
## 2: 2.18e-14
```

```
## 3: 4.44e-12
## 4: 1.86e-09
## 5: 1.08e-06
## 6: 1.98e-05
```



Annotation Cluster 1	Enrichment Score: 9.18	G		Count	P_Value	Benjamini
GOTERM_BP_DIRECT	canonical glycolysis	RT		14	1.7E-19	1.7E-16
GOTERM_BP_DIRECT	glycolytic process	RT		14	1.4E-17	6.8E-15
UP_KEYWORDS	Glycolysis	RT		12	3.5E-15	8.3E-13
KEGG_PATHWAY	Glycolysis / Gluconeogenesis	RT		14	1.6E-12	2.2E-10
BIOCARTA	Glycolysis Pathway	RT		7	1.1E-9	6.7E-8
KEGG_PATHWAY	Carbon metabolism	RT		13	1.7E-8	7.9E-7
GOTERM_BP_DIRECT	gluconeogenesis	RT		9	2.1E-8	6.8E-6
KEGG_PATHWAY	Biosynthesis of amino acids	RT		11	2.7E-8	9.7E-7
KEGG_PATHWAY	Biosynthesis of antibiotics	RT		16	6.3E-8	1.8E-6
KEGG_PATHWAY	Metabolic pathways	RT		32	8.0E-5	1.4E-3
UP_SEQ_FEATURE	binding site:Substrate	RT		12	1.7E-4	2.8E-2
UP_SEQ_FEATURE	active site:Proton acceptor	RT		12	6.2E-2	9.4E-1

Figure 3: Example: Annotation Cluster 1

```
# functional.annotaton.clustering.dat <- read.csv(textConnection(myfile), header=F)
# head(functional.annotaton.clustering.dat)
```

Current Gene List: List_1 Current Background: Homo sapiens 213 DAVID IDs ——— Importing the Functional Annotation Chart

```
URL2 <- "https://david.ncifcrf.gov/data/download/chart_0C32396F8F541508972594179.txt"
Functional_Annotation_Chart <- fread(URL2, header=T)
head(Functional_Annotation_Chart)
```

```
##            Category                                       Term Count    %
## 1: GOTERM_BP_DIRECT          GO:0061621~canonical glycolysis    14 6.57
## 2: GOTERM_BP_DIRECT            GO:0006096~glycolytic process    14 6.57
## 3:      UP_KEYWORDS                                Glycolysis    12 5.63
## 4:     KEGG_PATHWAY    hsa00010:Glycolysis / Gluconeogenesis    14 6.57
## 5:     KEGG_PATHWAY hsa00051:Fructose and mannose metabolism    10 4.69
## 6:         BIOCARTA   h_glycolysisPathway:Glycolysis Pathway     7 3.29
##      PValue
## 1: 1.68e-19
## 2: 1.38e-17
## 3: 3.48e-15
## 4: 1.57e-12
## 5: 1.42e-10
## 6: 1.07e-09
##                                                                                 Genes
## 1: ALDOA, PFKL, PFKFB4, PFKFB3, ALDOC, HK2, PGAM1, PFKP, HK1, TPI1, ENO2, PGK1, GAPDH, ENO1
## 2:  ALDOA, LDHA, PFKL, ALDOC, HK2, PGAM1, HK1, TPI1, PGM1, ENO2, PGK1, GAPDH, EDARADD, ENO1
## 3:                 ALDOA, TPI1, PFKL, ALDOC, ENO2, PGAM1, HK2, PFKP, HK1, PGK1, GAPDH, ENO1
## 4:     ALDOA, LDHA, PFKL, ALDOC, HK2, PGAM1, PFKP, HK1, TPI1, PGM1, ENO2, PGK1, GAPDH, ENO1
## 5:                           ALDOA, TPI1, MPI, PFKL, PFKFB4, PFKFB3, ALDOC, HK2, PFKP, HK1
## 6:                                                TPI1, PFKL, PGAM1, HK1, PGK1, GAPDH, ENO1
##    List Total Pop Hits Pop Total Fold Enrichment Bonferroni Benjamini
## 1:        183       26     16792           49.4   1.65e-16  1.65e-16
```

```
## 2:        183       34     16792            37.8   1.36e-14  6.80e-15
## 3:        199       31     20581            40.0   8.29e-13  8.29e-13
## 4:         90       67      6910            16.0   2.22e-10  2.22e-10
## 5:         90       32      6910            24.0   2.01e-08  1.00e-08
## 6:         25       10      1625            45.5   6.66e-08  6.66e-08
##         FDR
## 1: 2.65e-16
## 2: 2.18e-14
## 3: 4.44e-12
## 4: 1.86e-09
## 5: 1.68e-07
## 6: 1.08e-06
```

Functional Annotation Table Current Gene List: List_1 Current Background: Homo sapiens 28387 DAVID IDs ——

```
#importing functional annotation table
URL3 <- "https://david.ncifcrf.gov/data/download/tr_0C32396F8F541508973016208.txt"
myfile <- getURL(URL3, ssl.verifyhost=FALSE, ssl.verifypeer=FALSE)
Functional_Annotation_Table <- read.delim(textConnection(myfile), header=T, sep = "\t")
```

```
head(Functional_Annotation_Table)
```

```
##        ID
## 1    GBE1
## 2  PFKFB3
## 3  PFKFB4
## 4 ADPRHL2
## 5   ABCB6
## 6   BSPRY
##                                                                  Gene.Name
## 1                                1,4-alpha-glucan branching enzyme 1(GBE1)
## 2           6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3(PFKFB3)
## 3           6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4(PFKFB4)
## 4                                    ADP-ribosylhydrolase like 2(ADPRHL2)
## 5 ATP binding cassette subfamily B member 6 (Langereis blood group)(ABCB6)
## 6                                  B-box and SPRY domain containing(BSPRY)
##        Species BBID BIOCARTA
## 1 Homo sapiens
## 2 Homo sapiens
## 3 Homo sapiens
## 4 Homo sapiens
## 5 Homo sapiens
## 6 Homo sapiens
##                                                  COG_ONTOLOGY
## 1                       Carbohydrate transport and metabolism,
## 2
## 3
## 4
## 5 Posttranslational modification, protein turnover, chaperones,
## 6
##
## 1
## 2                   GO:0006000~fructose metabolic process,GO:0006003~fructose 2,6-bisphosphate metal
## 3                                          GO:0006000~fructose metabolic process,GO:0006003~fi
```

```
## 4
## 5 GO:0006779~porphyrin-containing compound biosynthetic process,GO:0006810~transport,GO:0006879~cellu
## 6
##
## 1
## 2
## 3
## 4
## 5 GO:0000139~Golgi membrane,GO:0005739~mitochondrion,GO:0005740~mitochondrial envelope,GO:0005741~mi
## 6
##
## 1                                                                                                GO:0003844~1,4-al
## 2
## 3
## 4
## 5 GO:0005524~ATP binding,GO:0015232~heme transporter activity,GO:0015439~heme-transporting ATPase act
## 6
##
## 1 IPR004193:Glycoside hydrolase, family 13, N-terminal,IPR006047:Glycosyl hydrolase, family 13, catal
## 2
## 3
## 4
## 5
## 6
##                                                                                                      KEGG_
## 1                                          hsa00500:Starch and sucrose metabolism,hsa01100:Metabolic pa
## 2 hsa00051:Fructose and mannose metabolism,hsa04066:HIF-1 signaling pathway,hsa04152:AMPK signaling p
## 3                                    hsa00051:Fructose and mannose metabolism,hsa04152:AMPK signaling p
## 4
## 5                                                                                 hsa02010:ABC transp
## 6
##
## 1
## 2
## 3
## 4
## 5 111600~Blood group, Langereis system,609153~Pseudohyperkalemia, familial, 2, due to red cell leak,
## 6
##                                    PIR_SUPERFAMILY                      SMART
## 1 PIRSF000463:1,4-alpha-glucan branching enzyme,              SM00642:Aamy,
## 2                                                           SM00855:SM00855,
## 3                                                           SM00855:SM00855,
## 4
## 5                                                               SM00382:AAA,
## 6                                                  SM00449:SPRY,SM00589:PRY,
##
## 1
## 2
## 3
## 4
## 5 3D-structure,Alternative splicing,ATP-binding,Cell membrane,Complete proteome,Disease mutation,Dysl
## 6
##
## 1
```

```
## 2 active site:Proton donor,active site:Tele-phosphohistidine intermediate,binding site:Fructose-6-pho
## 3                                  active site:Proton donor,active site:Tele-phosphohistidine interm
## 4
## 5
## 6
```