

Homework 7: Enrichment Analysis in R

Aizhan Uteubayeva (netID:au198)

2023-10-11

R Markdown - Enrichment analysis using EnrichR R package

Summary - Input gene lists created from T-Test / EdgeR group comparison analysis . Or an short list of genes of interest from your collaborator (Researcher/Clinician/) - If needed, short list genes based on p-value or FDR value. - Run EnrichR on short listed gene list

Import T-test results for further shortlisting

```
#folder that contains group comparison results
fileName <- "input/TTest_HW_EGradeIII-IV_(Comp).vs._EGradeI-II_(Base).TTest.csv"

# read in group comparison results
result1 <- read.csv(file = fileName,
                    header = T, stringsAsFactors = F, row.names = 1)
```

Filtering/shortlisting

Short list results based in p-value cut off (p-value ≤ 0.01), and fold change cut off (greater than 1.5 or less than -1.5) -à refer to R code done in class

```
pValueCutOff <- 0.01
which1 <- ((as.numeric(result1$Pvalue) <= pValueCutOff) & (as.numeric(result1$SignedFC) >= 1.5 | as.numeric(result1$SignedFC) <= -1.5))
table(which1) #82 genes TRUE

## which1
## FALSE TRUE
## 41845 82

resultShort <- result1[which1, ] #short listed genes
```

Export the short listed results for reference

```
# length of short listed results
nrow(resultShort)

## [1] 82

write.csv(x = resultShort, file = "output/HCC_TTest_results_shortlist2.csv")
```

Clean gene names

Genes are in the name contain “|”

```

funcSplit <- function(rep_gene) {
  rep_gene_split <- unlist(strsplit(x = rep_gene,
                                   split = "|",
                                   fixed = TRUE))

  gene <- rep_gene_split[2]
  return(gene)
}
geneListSplit <- apply(X = as.matrix(row.names(resultShort)),
                      MARGIN = 1, FUN = funcSplit )
head(geneListSplit) #cleaned gene names

## [1] "MYLK"      "FZD6"      "SRSF4"     "GMDS-DT"  "POMZP3"   "HINT1"

#remove duplicates
geneListSplit1 <- unique(geneListSplit)

# remove NA value
geneListFinal <- na.omit(geneListSplit1)

#print number of unique genes
length(geneListFinal)

## [1] 74

write.table(x = geneListFinal,
           file = "output/shortListedUniqueGenes.tsv",
           quote = F, sep = "\t", row.names = FALSE, col.names = F)

```

Loading Databases for EnrichrR package

```

#checking if Enrichr website and packing are working
#dbs <- enrichR::listEnrichrDbs() #total number of databases available = 200+

#testing if Enrichr package is working
testOutput <- enrichR::enrichr(genes = c("Runx1", "Gfi1", "Gfi1b", "Spi1", "Gata1", "Kdr"), databases =

## Uploading data to Enrichr... Done.
## Querying KEGG_2021_Human... Done.
## Parsing results... Done.

head(testOutput[[1]])

```

		Term	Overlap	P.value	Adjusted.P.value
## 1		Acute myeloid leukemia	2/67	0.0001643951	0.002794717
## 2	Transcriptional misregulation in cancer		2/192	0.0013407651	0.011396503
## 3	Pathways in cancer		2/531	0.0098313553	0.055711013
## 4	VEGF signaling pathway		1/59	0.0175720140	0.074681059
## 5	Chronic myeloid leukemia		1/76	0.0225871297	0.076796241
## 6	Th17 cell differentiation		1/107	0.0316774286	0.079200112
##	Old.P.value	Old.Adjusted.P.value	Odds.Ratio	Combined.Score	Genes
## 1	0	0	153.30000	1335.73937	SPI1;RUNX1
## 2	0	0	52.11579	344.72067	SPI1;RUNX1
## 3	0	0	18.39792	85.03847	SPI1;RUNX1
## 4	0	0	68.74483	277.82863	KDR
## 5	0	0	53.11733	201.33461	RUNX1

```
## 6          0          0  37.52453    129.54033    RUNX1
```

```
#List of databases for which enrichment analysis will be run
dblist1 <- read.csv(file = "input/2023-EnrichR-Databases.txt",
                    header = F, stringsAsFactors = F)
```

```
head(dblist1)
```

```
##          V1
## 1      KEGG_2021_Human
## 2  WikiPathway_2021_Human
## 3 GO_Biological_Process_2023
## 4      Reactome_2022
## 5      BioPlanet_2019
## 6      ClinVar_2019
```

Calling function to run Enrichment

```
# set output file name
outputFileName1 <- paste("output/HCC", "_EnrichR.xlsx", sep="")
```

```
#Load R script into the environment
source(file = "input/functionEnrichment.R")
```

```
#call function to run Enrichment
functionEnrichment(dblist1, geneListFinal, outputFileName1)
```

```
## Uploading data to Enrichr... Done.
##   Querying KEGG_2021_Human... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
##   Querying WikiPathway_2021_Human... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
##   Querying GO_Biological_Process_2023... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
##   Querying Reactome_2022... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
##   Querying BioPlanet_2019... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
##   Querying ClinVar_2019... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
##   Querying Transcription_Factor_PPis... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
##   Querying TRANSFAC_and_JASPAR_PWMs... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
##   Querying TargetScan_microRNA... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
```

```

## Querying miRTarBase_2017... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying DisGeNET... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying OMIM_Disease... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying Jensen_DISEASES... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying Chromosome_Location... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying VirusMINT... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying Virus-Host_PPI_P-HIPSTer_2020... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying HMDB_Metabolites... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying dbGap... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying MSigDB_Hallmark_2020... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying ProteomicsDB_2020... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying GWAS_Catalog_2023... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying InterPro_Domains_2019... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying CCLE_Proteomics_2020... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying Proteomics_Drug_Atlas_2023... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying PheWeb_2019... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying Pfam_Domains_2019... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.
## Querying ChEA_2022... Done.
## Parsing results... Done.
## Uploading data to Enrichr... Done.

```

```
## Querying SILAC_Phosphoproteomics... Done.  
## Parsing results... Done.  
## Uploading data to Enrichr... Done.  
## Querying Azimuth_2023... Done.  
## Parsing results... Done.  
## Uploading data to Enrichr... Done.  
## Querying MAGNET_2023... Done.  
## Parsing results... Done.  
## Uploading data to Enrichr... Done.  
## Querying GeDiPNet_2023... Done.  
## Parsing results... Done.
```