Functional Pathways Analysis

2022-04-22

Tutorial links:
https://github.com/twbattaglia/RNAseq-workflow
http://yulab-smu.top/biomedical-knowledge-mining-book/clusterprofiler-go.html

Install Packages

Open Rstudio and Install Bioconductor and other packages

```
if (!require("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install(version = "3.14")
BiocManager::install("clusterProfiler")
BiocManager::install("AnnotationHub")
BiocManager::install("org.Hs.eg.db")
BiocManager::install("pathview")
```

Load libraries

```
library(clusterProfiler)
library(AnnotationHub)
library(org.Hs.eg.db)
library(DESeq2)
library(pathview)
```

Check the organism in annotation hub

```
hub <- AnnotationHub()
#If throws corrupt message rstart R session

unique(hub$dataprovider)
head(unique(hub$species))
```

load the dds object and get your desired result

```
dds <- readRDS("Analysis/dds.rds")
resultsNames(dds)
# The list here is just an example
ressults <- results(dds, contrast = list( c("classCHLA01_211", "classD425_211", "classDAOY_211"),
                                    c("classCHLA01_vo", "classD425_vo", "classDAOY_vo" )),
                                  listValues = c(1/3, -1/3))
```

Use mapIds to convert the gene name to Entrez ID or ENSEMBL ID

```
results$entrez <- mapIds(x = org.Hs.eg.db,
                         keys = row.names(results),
                         column = "ENTREZID",
                         keytype = "SYMBOL",
                         multiVals = "first")

# Add ENSEMBL
results$ensembl <- mapIds(x = org.Hs.eg.db,
                          keys = row.names(results),
                          column = "ENSEMBL",
                          keytype = "SYMBOL",
                          multiVals = "first")
```

Subset for only significant genes

```
results_sig <- subset(results, padj < 0.05)
# You can use logfold change in filter also
#results_sig <- subset(results, padj < 0.05 & abs(log2FoldChange) > 0.5)

# Remove any genes that do not have any entrez identifiers
results_sig_entrez <- subset(results_sig, is.na(entrez) == FALSE)
results_sig_ensembl <- subset(results_sig, is.na(ensembl) == FALSE)
# Create a matrix of gene log2 fold changes
gene_matrix <- results_sig_entrez$log2FoldChange
gene_matrix2 <- results_sig_ensembl$log2FoldChange
# Add the entrezID's as names for each logFC entry
names(gene_matrix) <- results_sig_entrez$entrez
names(gene_matrix2) <- results_sig_ensembl$ensembl
```

Sse functions from *clusterProfiler* to get enrich pathways

```
# Enrich genes using the KEGG database
kegg_enrich <- enrichKEGG(gene = names(gene_matrix),
                          organism = 'hsa',
                          pvalueCutoff = 0.05,
                          qvalueCutoff = 0.10)
# Plot results
barplot(kegg_enrich,
        drop = TRUE,
        showCategory = 15,
        title = "KEGG Enrichment Pathways",
        font.size = 8)
ggsave("KEGG_enrich_pathways.pdf", height = 5, width = 5)

# Enrich genes using the Gene Onotlogy
#GO_Biological_pathways
go_enrich <- enrichGO(gene = names(gene_matrix),
                      OrgDb = 'org.Hs.eg.db',
                      readable = T,
                      keyType = 'ENTREZID', #ENSEMBL,
                      ont = "BP",
                      pAdjustMethod = "BH",
                      pvalueCutoff = 0.05,
                      qvalueCutoff = 0.10)
# Plot results
barplot(go_enrich,
        drop = TRUE,
        showCategory = 15,
        title = "GO Biological Pathways",
        font.size = 8)
ggsave("GO_Biological_pathways.pdf", height = 5, width = 5)

#GO_CelluarComponet_pathways
go_enrich <- enrichGO(gene = names(gene_matrix),
                      OrgDb = 'org.Hs.eg.db',
                      readable = T,
                      keyType = 'ENTREZID', #ENSEMBL,
                      ont = "CC",
                      pAdjustMethod = "BH",
                      pvalueCutoff = 0.05,
                      qvalueCutoff = 0.10)
# Plot results
barplot(go_enrich,
        drop = TRUE,
        showCategory = 15,
        title = "GO CC Pathways",
        font.size = 8)
ggsave("Analysis/plots/MIR211_vs_VO_enrichment/GO_CelluarComponet_pathways.pdf", height = 5, width = 5)

#GO_MolecularFunction_pathways
go_enrich <- enrichGO(gene = names(gene_matrix),
                      OrgDb = 'org.Hs.eg.db',
                      readable = T,
                      keyType = 'ENTREZID', #ENSEMBL,
                      ont = "MF",
                      pAdjustMethod = "BH",
                      pvalueCutoff = 0.05,
                      qvalueCutoff = 0.10)
# Plot results
barplot(go_enrich,
        drop = TRUE,
        showCategory = 15,
        title = "GO MF Pathways",
        font.size = 8)
ggsave("GO_MolecularFunction_pathways.pdf", height = 5, width = 5)
```

Plotting KEGG Pathways

```
# Pathview is a package that can take KEGG identifier and overlay fold
# changes to the genes which are found to be significantly
# different. Pathview also works with other organisms found in
# the KEGG database and can plot any of the KEGG pathways for
# the particular organism.

# Plot specific KEGG pathways (with fold change)
# pathway.id : KEGG pathway identifier
# KEEG pathway IDs: https://www.genome.jp/kegg/pathway.html
pathview(gene.data = gene_matrix,
         pathway.id = c("05200", "05206"),  # c(pathways in cancer, microRNA in cancer)
         species = "hsa")
```