GSEA

Hildah Njoroge

2023-08-30

Step 1

Load the SigRes data

```
#HTseq data
load("C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/scripts/DEG_analysis/Gene_counts/hisat2_htseq_pipeline/deseq2_statistical_analysis_HTSeq.Rdata")

#HTG data
load("C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/scripts/DEG_analysis/Gene_counts/htg_pipeline/deseq2_statistical_analysis_HTG.Rdata")
```

Step 2

Install and load the necessary packages

```
if (!require(clusterProfiler, quietly = T)) {install.packages("clusterProfiler")}
```

```
## Warning: package 'clusterProfiler' was built under R version 4.3.1
```

```
##
```

```
## Registered S3 methods overwritten by 'treeio':
##   method              from
##   MRCA.phylo          tidytree
##   MRCA.treedata       tidytree
##   Nnode.treedata      tidytree
##   Ntip.treedata       tidytree
##   ancestor.phylo      tidytree
##   ancestor.treedata   tidytree
##   child.phylo         tidytree
##   child.treedata      tidytree
##   full_join.phylo     tidytree
##   full_join.treedata  tidytree
##   groupClade.phylo    tidytree
##   groupClade.treedata tidytree
##   groupOTU.phylo      tidytree
##   groupOTU.treedata   tidytree
##   inner_join.phylo    tidytree
##   inner_join.treedata tidytree
```

```
##   is.rooted.treedata  tidytree
##   nodeid.phylo        tidytree
##   nodeid.treedata     tidytree
##   nodelab.phylo       tidytree
##   nodelab.treedata    tidytree
##   offspring.phylo     tidytree
##   offspring.treedata  tidytree
##   parent.phylo        tidytree
##   parent.treedata     tidytree
##   root.treedata       tidytree
##   rootnode.phylo      tidytree
##   sibling.phylo       tidytree
```

```
## clusterProfiler v4.8.2  For help: https://yulab-smu.top/biomedical-knowledge-mining-book/
##
## If you use clusterProfiler in published research, please cite:
## T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo,
and G Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The In
novation. 2021, 2(3):100141
```

```
##
## Attaching package: 'clusterProfiler'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(clusterProfiler)
if (!require(org.Mm.eg.db, quietly = T)) {BiocManager::install("org.Mm.eg.db")}
```

```
## Warning: package 'AnnotationDbi' was built under R version 4.3.1
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##     table, tapply, union, unique, unsplit, which.max, which.min
```

```
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

```
##
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:clusterProfiler':
##
##     rename
```

```
## The following object is masked from 'package:utils':
##
##     findMatches
```

```
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname
```

```
##
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:clusterProfiler':
##
##     slice
```

```
## The following object is masked from 'package:grDevices':
##
##     windows
```

```
##
## Attaching package: 'AnnotationDbi'
```

```
## The following object is masked from 'package:clusterProfiler':
##
##     select
```

```
##
```

```
library(org.Mm.eg.db)
library(org.Hs.eg.db)
```

```
##
```

Step 3

Run GO-SEA

A. WT vs colOnly

HTSeq

```
#GO SEA
goSEA_HTseq <- enrichGO(
  gene = rownames(sigRes_HTSeq),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)
```

Step 4 Visualisation

```
# cnetplot(goSEA, colorEdge = TRUE, cex_label_gene = 0.9)
# dotplot(goSEA)
#
# cnetplot(goSEA_HTseq, colorEdge = TRUE, cex_label_gene = 0.5)
# dotplot(goSEA_HTseq)
# goSEA_HTseq[1:10,1:7]
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTseq)
paste("Number of pathways differentially regulated is", nrow(goSEA_HTseq))
```

```
## [1] "Number of pathways differentially regulated is 2593"
```

```
# goSEA_HTseq[1:10,1:7]
results_goSEA_HTseq <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjus
t", "qvalue"))
results_goSEA_HTseq <- t(results_goSEA_HTseq)
colnames(results_goSEA_HTseq) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust
", "qvalue"); results_goSEA_HTseq <- results_goSEA_HTseq[-1,]
results_goSEA_HTseq <- goSEA_HTseq[1:10,1:7]
```

HTG WT vs colOnly

```
#GO SEA
goSEA_HTG <- enrichGO(
```

```
  gene = rownames(sigRes_HTG),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)
```

Step 4 Visualisation

```
# cnetplot(goSEA, colorEdge = TRUE, cex_label_gene = 0.9)
# dotplot(goSEA)


# cnetplot(goSEA_HTG, colorEdge = T, cex_label_gene = 0.5)
# dotplot(goSEA_HTG)
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTG)
paste("WT vs ColOnly, HTG - Number of pathways differentially regulated is", nrow(goSEA_HTG))
```

```
## [1] "WT vs ColOnly, HTG - Number of pathways differentially regulated is 2563"
```

```
paste("WT vs ColOnly, HTSeq - Number of pathways differentially regulated is", nrow(goSEA_HTseq))
```

```
## [1] "WT vs ColOnly, HTSeq - Number of pathways differentially regulated is 2593"
```

```
# goSEA_HTG[1:10,1:7]
results_goSEA_HTG <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust", "qvalue"))
results_goSEA_HTG <- t(results_goSEA_HTG)
colnames(results_goSEA_HTG) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust", "qvalue"); results_goSEA_HTG <- results_goSEA_HTG[-1,]
results_goSEA_HTG <- goSEA_HTG[1:10,1:7]
```

```
results1 <- data.frame(top10pathways_HTG=results_goSEA_HTG[,1],top10pathways_HTSeq=results_goSEA_HTseq[,1])
results1
```

```
##    top10pathways_HTG top10pathways_HTSeq
## 1         GO:0007159          GO:0019221
## 2         GO:0002443          GO:0007159
## 3         GO:1903037          GO:1903037
```

```
## 4         GO:0019221          GO:0050863
## 5         GO:0002460          GO:0002460
## 6         GO:0050863          GO:0070663
## 7         GO:0050900          GO:0070661
## 8         GO:0002697          GO:0002443
## 9         GO:0070663          GO:0050900
## 10        GO:0002696          GO:0032944
```

```
write.csv(results1,file = "C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/results/pipeline_comparisons/top10_diff_reg_pathways_nocol_colonly.csv")
setdiff(results_goSEA_HTG[,2],results_goSEA_HTseq[,2]) ## meaning what is in the first that isn't in the second set
```

```
## [1] "regulation of immune effector process"
## [2] "positive regulation of leukocyte activation"
```

```
setdiff(results_goSEA_HTseq[,2],results_goSEA_HTG[,2])
```

```
## [1] "leukocyte proliferation"
## [2] "regulation of mononuclear cell proliferation"
```

ggplot for visualisation

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
library(ggrepel)
library(stringr)
# df <- goSEA_HTG@result[1:10,]
df <- results_goSEA_HTG[1:10,]
head(df)
```

```
##                    ID
## GO:0007159 GO:0007159
## GO:0002443 GO:0002443
## GO:1903037 GO:1903037
## GO:0019221 GO:0019221
## GO:0002460 GO:0002460
## GO:0050863 GO:0050863
##
                              Description
## GO:0007159
             leukocyte cell-cell adhesion
## GO:0002443
              leukocyte mediated immunity
## GO:1903037                                                                r
egulation of leukocyte cell-cell adhesion
```

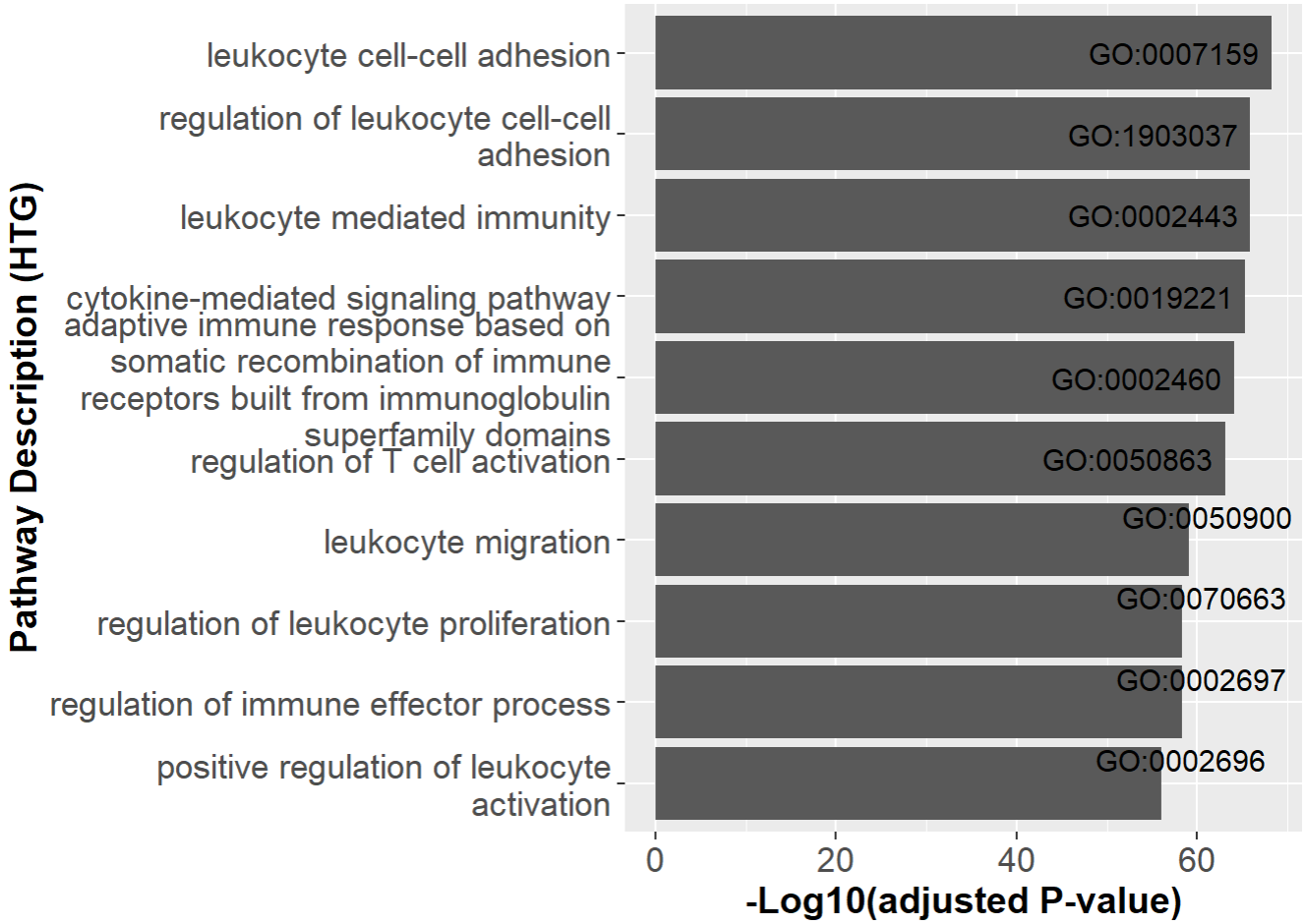
```
## GO:0019221
      cytokine-mediated signaling pathway
## GO:0002460 adaptive immune response based on somatic recombination of immune receptors buil
t from immunoglobulin superfamily domains
## GO:0050863
          regulation of T cell activation
##             GeneRatio   BgRatio       pvalue     p.adjust       qvalue
## GO:0007159    86/423 420/28564 1.315383e-72 6.104691e-69 2.571227e-69
## GO:0002443    88/423 480/28564 6.475255e-70 1.370215e-66 5.771191e-67
## GO:1903037    81/423 380/28564 8.857240e-70 1.370215e-66 5.771191e-67
## GO:0019221    84/423 429/28564 3.762650e-69 4.365614e-66 1.838748e-66
## GO:0002460    79/423 373/28564 8.778517e-68 8.148219e-65 3.431938e-65
## GO:0050863    79/423 384/28564 1.005391e-66 7.776702e-64 3.275459e-64
```

```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                                "Description"
## [3] "GeneRatio"                         "BgRatio"
## [5] "pvalue"                            "p.adjust"
## [7] "qvalue"                            "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df,
       aes(x = Negative_log10_adjusted_PValue,
           y = reorder(Pathway_Description,
                       Negative_log10_adjusted_PValue),
           label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(colour="black",
    size=4,
    min.segment.length = 10,
    hjust = 0.5,
    vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)",
       y = "Pathway Description (HTG)") +
  theme(axis.text=element_text(size=13),
        axis.title=element_text(size=14,face="bold"))
```




```
df <- results_goSEA_HTseq[1:10,]
head(df)
```

```
##                    ID
## GO:0019221 GO:0019221
## GO:0007159 GO:0007159
## GO:1903037 GO:1903037
## GO:0050863 GO:0050863
## GO:0002460 GO:0002460
## GO:0070663 GO:0070663
##
                             Description
## GO:0019221
      cytokine-mediated signaling pathway
## GO:0007159
             leukocyte cell-cell adhesion
## GO:1903037                                                                    r
egulation of leukocyte cell-cell adhesion
## GO:0050863
          regulation of T cell activation
## GO:0002460 adaptive immune response based on somatic recombination of immune receptors buil
t from immunoglobulin superfamily domains
## GO:0070663
    regulation of leukocyte proliferation
##            GeneRatio   BgRatio       pvalue     p.adjust       qvalue
```

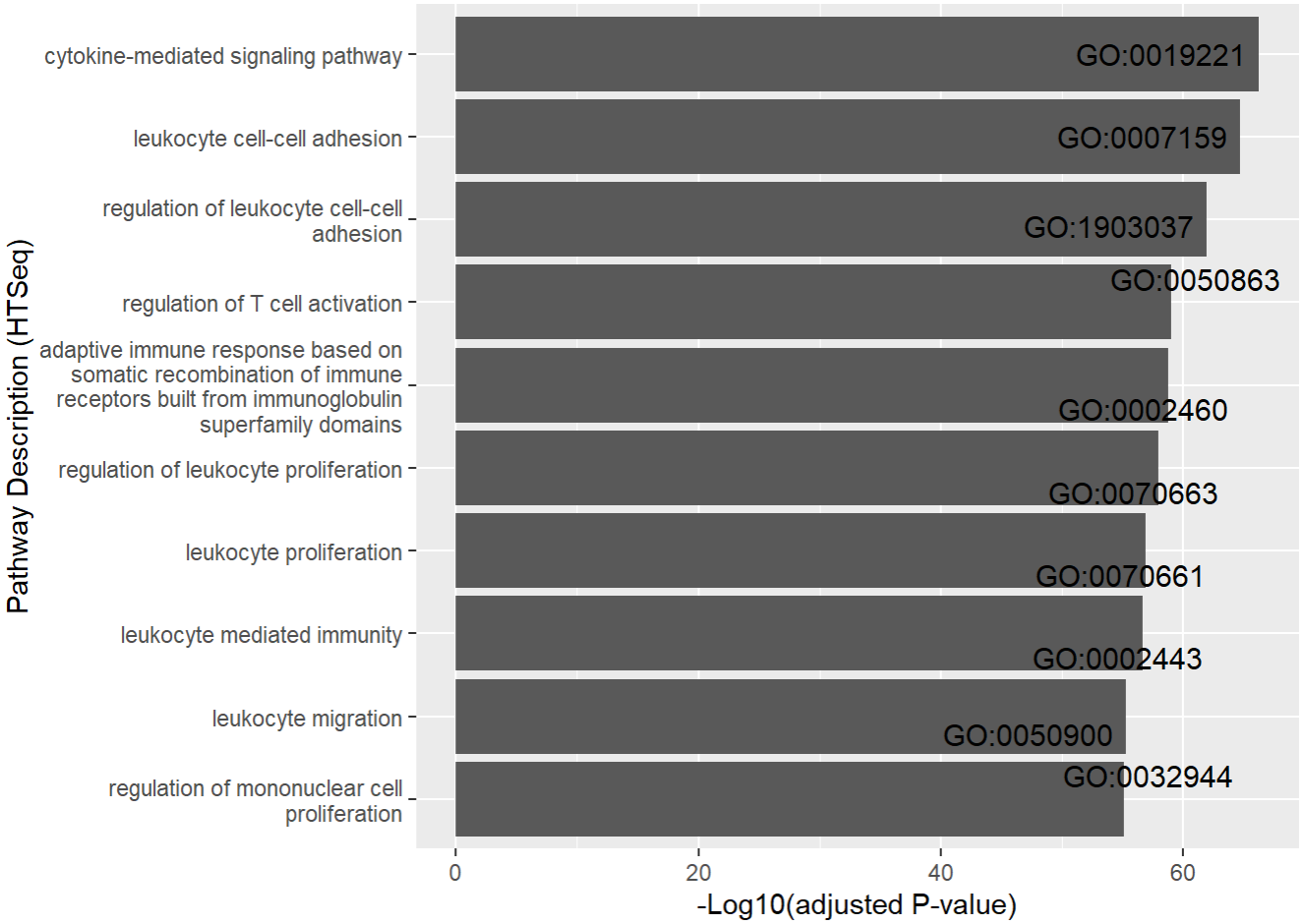
```
## GO:0019221    84/407 429/28564 1.231580e-70 5.678813e-67 2.368522e-67
## GO:0007159    82/407 420/28564 8.475137e-69 1.953943e-65 8.149513e-66
## GO:1903037    77/407 380/28564 8.180200e-66 1.257297e-62 5.243939e-63
## GO:0050863    75/407 384/28564 8.343827e-63 9.618347e-60 4.011624e-60
## GO:0002460    74/407 373/28564 1.761955e-62 1.624875e-59 6.777036e-60
## GO:0070663    67/407 287/28564 1.515175e-61 1.164412e-58 4.856535e-59
```

```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                                "Description"
## [3] "GeneRatio"                         "BgRatio"
## [5] "pvalue"                            "p.adjust"
## [7] "qvalue"                            "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTSeq)")
```



Step 3

Run GO-SEA

HTSeq WT vs Colnodysp

```
#GO SEA
goSEA_HTseq_Colnodysp <- enrichGO(
  gene = rownames(sigRes_HTSeq_Colnodysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)

goSEA_HTseq <- goSEA_HTseq_Colnodysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTseq)
paste("Number of pathways differentially regulated is", nrow(goSEA_HTseq))
```

```
## [1] "Number of pathways differentially regulated is 2456"
```

```
# goSEA_HTseq[1:10,1:7]
results_goSEA_HTseq <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjus
t", "qvalue"))
results_goSEA_HTseq <- t(results_goSEA_HTseq)
colnames(results_goSEA_HTseq) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust
", "qvalue"); results_goSEA_HTseq <- results_goSEA_HTseq[-1,]
results_goSEA_HTseq <- goSEA_HTseq[1:10,1:7]
```

HTG WT vs Colnodysp

```
#GO SEA
goSEA_HTG_Colnodysp <- enrichGO(
  gene = rownames(sigRes_HTG_Colnodysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)

goSEA_HTG <- goSEA_HTG_Colnodysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTG)
paste("WT vs Colitis without dysp, HTG - Number of pathways differentially regulated is", nrow
(goSEA_HTG))
```

```
## [1] "WT vs Colitis without dysp, HTG - Number of pathways differentially regulated is 2456"
```

```
paste("WT vs Colitis without dysp, HTSeq - Number of pathways differentially regulated is", nr
ow(goSEA_HTseq))
```

```
## [1] "WT vs Colitis without dysp, HTSeq - Number of pathways differentially regulated is 245
6"
```

```
# goSEA_HTG[1:10,1:7]
results_goSEA_HTG <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust"
, "qvalue"))
results_goSEA_HTG <- t(results_goSEA_HTG)
colnames(results_goSEA_HTG) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust",
 "qvalue"); results_goSEA_HTG <- results_goSEA_HTG[-1,]
results_goSEA_HTG <- goSEA_HTG[1:10,1:7]
```

Then, compare the top 10 differentially regulated pathways in the nocol vs colnodysp groups

```
results2 <- data.frame(top10pathways_HTG=results_goSEA_HTG[,1],top10pathways_HTSeq=results_goS
EA_HTseq[,1])
results2
```

```
##    top10pathways_HTG top10pathways_HTSeq
## 1         GO:0002443          GO:0019221
## 2         GO:0007159          GO:0007159
## 3         GO:1903037          GO:1903037
## 4         GO:0002460          GO:0002460
## 5         GO:0019221          GO:0050863
## 6         GO:0050863          GO:0002443
## 7         GO:0002696          GO:0050900
## 8         GO:0050867          GO:0002696
## 9         GO:0002449          GO:0050867
## 10        GO:0070661          GO:0070661
```

```
write.csv(results2,file = "C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/
results/pipeline_comparisons/top10_diff_reg_pathways_nocol_colnodysp.csv")
setdiff(results_goSEA_HTG[,2],results_goSEA_HTseq[,2])
```

```
## [1] "lymphocyte mediated immunity"
```

```
setdiff(results_goSEA_HTseq[,2],results_goSEA_HTG[,2])
```

```
## [1] "leukocyte migration"
```

ggplot for visualisation

```
library(stringr)
# df <- goSEA_HTG@result[1:10,]
df <- results_goSEA_HTG[1:10,]
head(df)
```

```
##                    ID
## GO:0002443 GO:0002443
## GO:0007159 GO:0007159
## GO:1903037 GO:1903037
## GO:0002460 GO:0002460
## GO:0019221 GO:0019221
## GO:0050863 GO:0050863
##
                              Description
## GO:0002443
              leukocyte mediated immunity
```

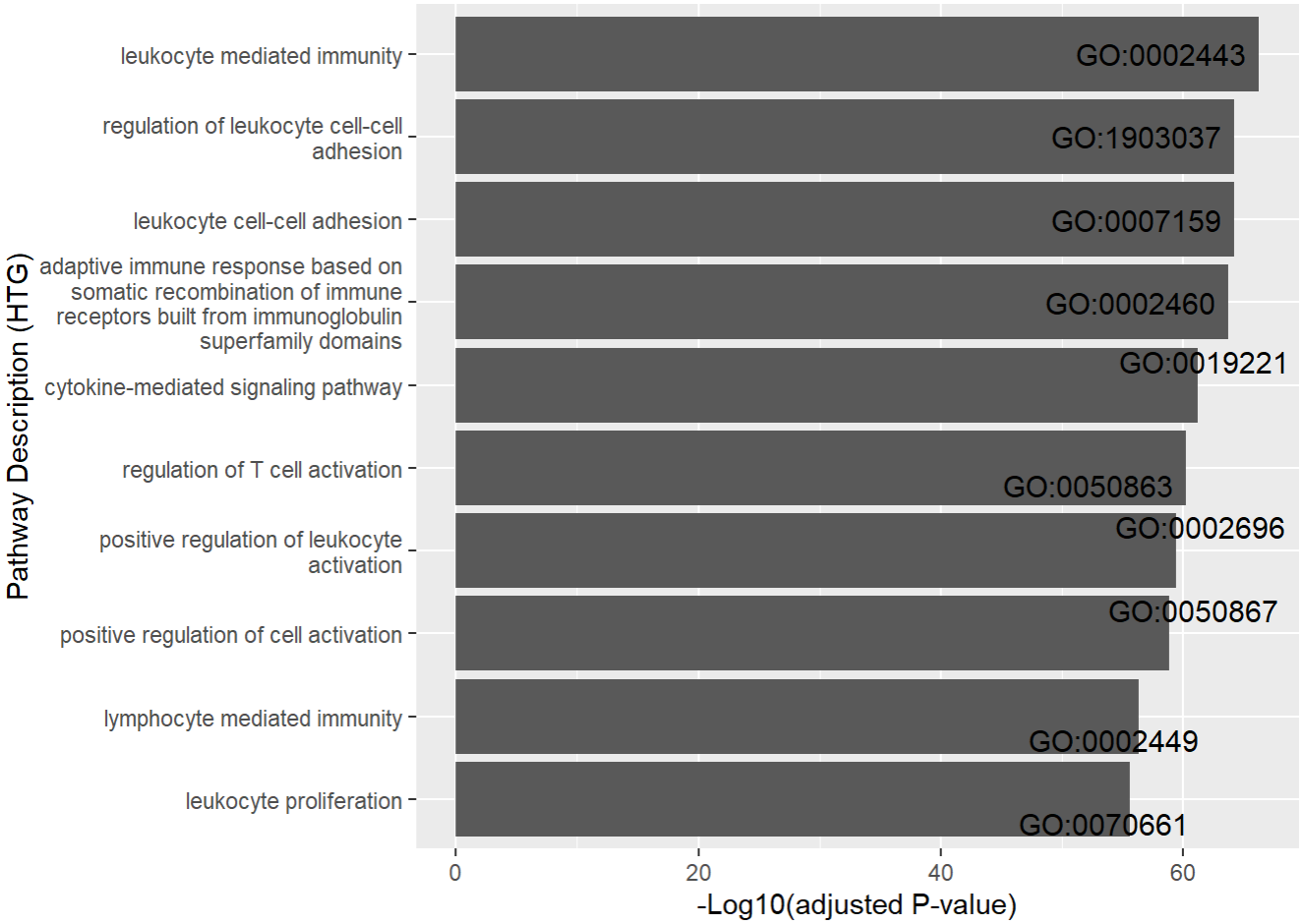
```
## GO:0007159
             leukocyte cell-cell adhesion
## GO:1903037                                                                      r
egulation of leukocyte cell-cell adhesion
## GO:0002460 adaptive immune response based on somatic recombination of immune receptors buil
t from immunoglobulin superfamily domains
## GO:0019221
      cytokine-mediated signaling pathway
## GO:0050863
          regulation of T cell activation
##             GeneRatio   BgRatio       pvalue     p.adjust       qvalue
## GO:0002443    90/442 480/28564 1.158186e-70 5.356609e-67 2.371233e-67
## GO:0007159    84/442 420/28564 2.776921e-68 5.722181e-65 2.533062e-65
## GO:1903037    81/442 380/28564 3.711685e-68 5.722181e-65 2.533062e-65
## GO:0002460    80/442 373/28564 1.587402e-67 1.835434e-64 8.124992e-65
## GO:0019221    82/442 429/28564 6.246960e-65 5.778438e-62 2.557966e-62
## GO:0050863    78/442 384/28564 7.441545e-64 5.736191e-61 2.539264e-61
```

```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                                "Description"
## [3] "GeneRatio"                         "BgRatio"
## [5] "pvalue"                            "p.adjust"
## [7] "qvalue"                            "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTG)")
```




```
df <- results_goSEA_HTseq[1:10,]
head(df)
```

```
##                    ID
## GO:0019221 GO:0019221
## GO:0007159 GO:0007159
## GO:1903037 GO:1903037
## GO:0002460 GO:0002460
## GO:0050863 GO:0050863
## GO:0002443 GO:0002443
##
                             Description
## GO:0019221
      cytokine-mediated signaling pathway
## GO:0007159
             leukocyte cell-cell adhesion
## GO:1903037                                                                    r
egulation of leukocyte cell-cell adhesion
## GO:0002460 adaptive immune response based on somatic recombination of immune receptors buil
t from immunoglobulin superfamily domains
## GO:0050863
          regulation of T cell activation
## GO:0002443
              leukocyte mediated immunity
##            GeneRatio   BgRatio       pvalue     p.adjust       qvalue
```

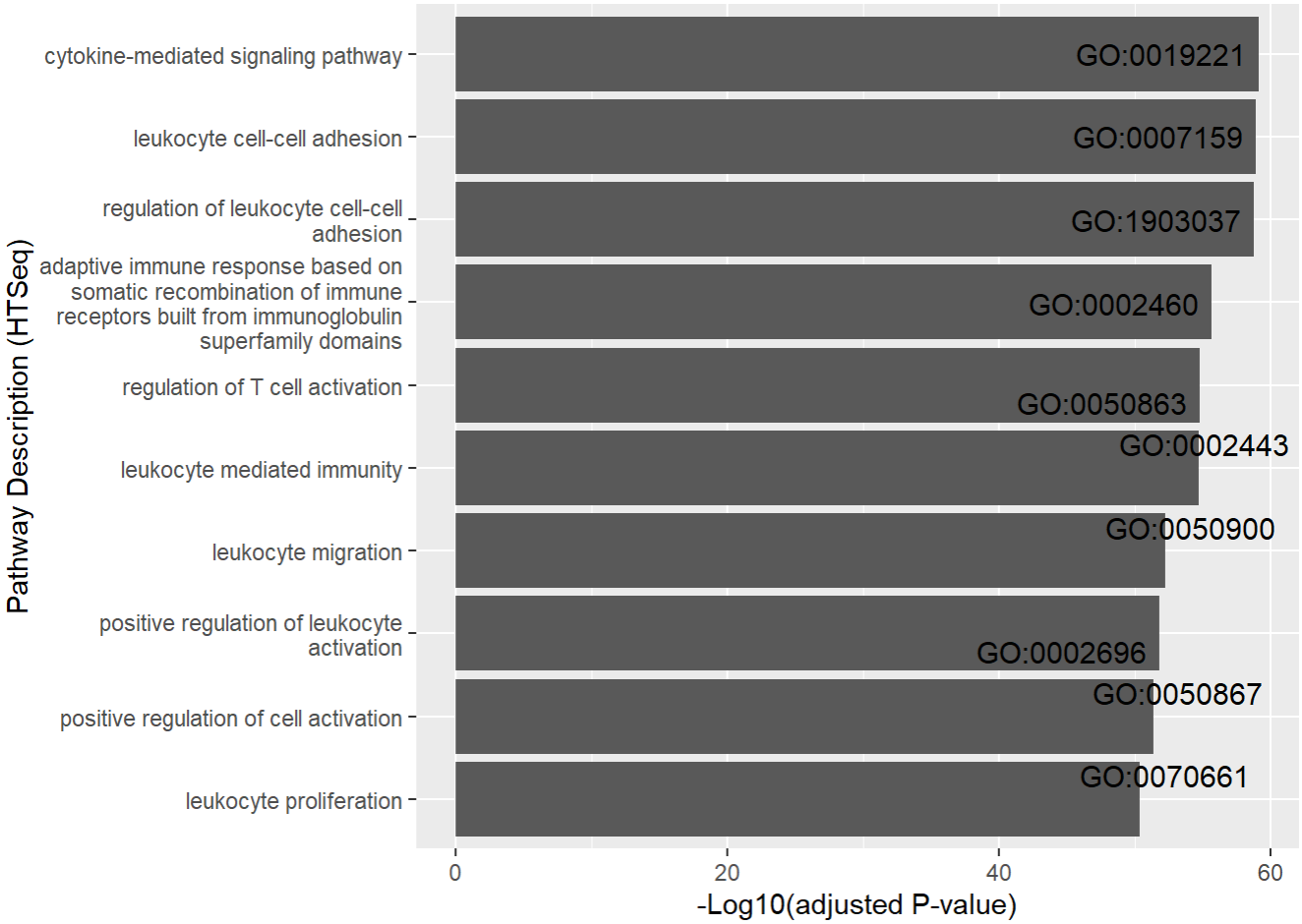
```
## GO:0019221    80/429 429/28564 1.700261e-63 7.858605e-60 3.466742e-60
## GO:0007159    79/429 420/28564 5.422014e-63 1.253027e-59 5.527600e-60
## GO:1903037    76/429 380/28564 1.134591e-62 1.748026e-59 7.711237e-60
## GO:0002460    73/429 373/28564 1.953418e-59 2.257174e-56 9.957291e-57
## GO:0050863    73/429 384/28564 1.791169e-58 1.655756e-55 7.304198e-56
## GO:0002443    79/429 480/28564 2.788608e-58 2.148158e-55 9.476376e-56
```

```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                                "Description"
## [3] "GeneRatio"                         "BgRatio"
## [5] "pvalue"                            "p.adjust"
## [7] "qvalue"                            "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTSeq)")
```



Step 3

Run GO-SEA

HTSeq WT vs Coldysp

```
#GO SEA
goSEA_HTseq_Coldysp <- enrichGO(
  gene = rownames(sigRes_HTSeq_Coldysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)


goSEA_HTseq <- goSEA_HTseq_Coldysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTseq)
```

```
paste("Number of pathways differentially regulated is", nrow(goSEA_HTseq))
```

```
## [1] "Number of pathways differentially regulated is 2733"
```

```
# goSEA_HTseq[1:10,1:7]
results_goSEA_HTseq <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjus
t", "qvalue"))
results_goSEA_HTseq <- t(results_goSEA_HTseq)
colnames(results_goSEA_HTseq) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust
", "qvalue"); results_goSEA_HTseq <- results_goSEA_HTseq[-1,]
results_goSEA_HTseq <- goSEA_HTseq[1:10,1:7]
```

HTG WT vs Coldysp

```
#GO SEA
goSEA_HTG_Coldysp <- enrichGO(
  gene = rownames(sigRes_HTG_Coldysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)

goSEA_HTG <- goSEA_HTG_Coldysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTG)
paste("WT vs Colitis with dysp, HTG - Number of pathways differentially regulated is", nrow(go
SEA_HTG))
```

```
## [1] "WT vs Colitis with dysp, HTG - Number of pathways differentially regulated is 2713"
```

```
paste("WT vs Colitis with dysp, HTSeq - Number of pathways differentially regulated is", nrow(
goSEA_HTseq))
```

```
## [1] "WT vs Colitis with dysp, HTSeq - Number of pathways differentially regulated is 2733"
```

```
# goSEA_HTG[1:10,1:7]
results_goSEA_HTG <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust"
, "qvalue"))
results_goSEA_HTG <- t(results_goSEA_HTG)
colnames(results_goSEA_HTG) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust",
 "qvalue"); results_goSEA_HTG <- results_goSEA_HTG[-1,]
results_goSEA_HTG <- goSEA_HTG[1:10,1:7]
```

Then, compare the top 10 differentially regulated pathways in the nocol vs colnodysp groups

```
results3 <- data.frame(top10pathways_HTG=results_goSEA_HTG[,2],top10pathways_HTSeq=results_goS
EA_HTseq[,2])
results3
```

```
##
                top10pathways_HTG
## 1
     leukocyte cell-cell adhesion
## 2
              leukocyte migration
## 3                                                                    regulatio
n of leukocyte cell-cell adhesion
## 4
  regulation of T cell activation
## 5                                                                          cy
tokine-mediated signaling pathway
## 6
      leukocyte mediated immunity
## 7                                                                      posit
ive regulation of cell activation
## 8                                                                  positive r
egulation of leukocyte activation
## 9  adaptive immune response based on somatic recombination of immune receptors built from i
mmunoglobulin superfamily domains
## 10                                                                       regu
lation of leukocyte proliferation
##
              top10pathways_HTSeq
## 1
     leukocyte cell-cell adhesion
## 2
              leukocyte migration
## 3                                                                          cy
tokine-mediated signaling pathway
## 4                                                                    regulatio
n of leukocyte cell-cell adhesion
## 5
  regulation of T cell activation
## 6
      leukocyte mediated immunity
## 7  adaptive immune response based on somatic recombination of immune receptors built from i
mmunoglobulin superfamily domains
## 8                                                                      posit
ive regulation of cell activation
## 9                                                                  positive r
egulation of leukocyte activation
## 10
          leukocyte proliferation
```

```
write.csv(results3,file = "C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/results/pipeline_comparisons/top10_diff_reg_pathways_nocol_coldysp.csv")
setdiff(results_goSEA_HTG[,2],results_goSEA_HTseq[,2])
```

```
## [1] "regulation of leukocyte proliferation"
```

```
setdiff(results_goSEA_HTseq[,2],results_goSEA_HTG[,2])
```

```
## [1] "leukocyte proliferation"
```

ggplot for visualisation

```
library(stringr)
# df <- goSEA_HTG@result[1:10,]
df <- results_goSEA_HTG[1:10,]
head(df)
```

```
##                    ID                                 Description GeneRatio
## GO:0007159 GO:0007159              leukocyte cell-cell adhesion    91/456
## GO:0050900 GO:0050900                       leukocyte migration    85/456
## GO:1903037 GO:1903037 regulation of leukocyte cell-cell adhesion    83/456
## GO:0050863 GO:0050863            regulation of T cell activation    81/456
## GO:0019221 GO:0019221        cytokine-mediated signaling pathway    83/456
## GO:0002443 GO:0002443                leukocyte mediated immunity    84/456
##               BgRatio       pvalue     p.adjust       qvalue
## GO:0007159 420/28564 3.439532e-76 1.645816e-72 6.625624e-73
## GO:0050900 386/28564 1.123336e-71 2.687581e-68 1.081950e-68
## GO:1903037 380/28564 1.225479e-69 1.954639e-66 7.868865e-67
## GO:0050863 384/28564 1.270678e-66 1.520049e-63 6.119320e-64
## GO:0019221 429/28564 4.967036e-65 4.753454e-62 1.913616e-62
## GO:0002443 480/28564 4.717898e-62 3.762524e-59 1.514694e-59
```

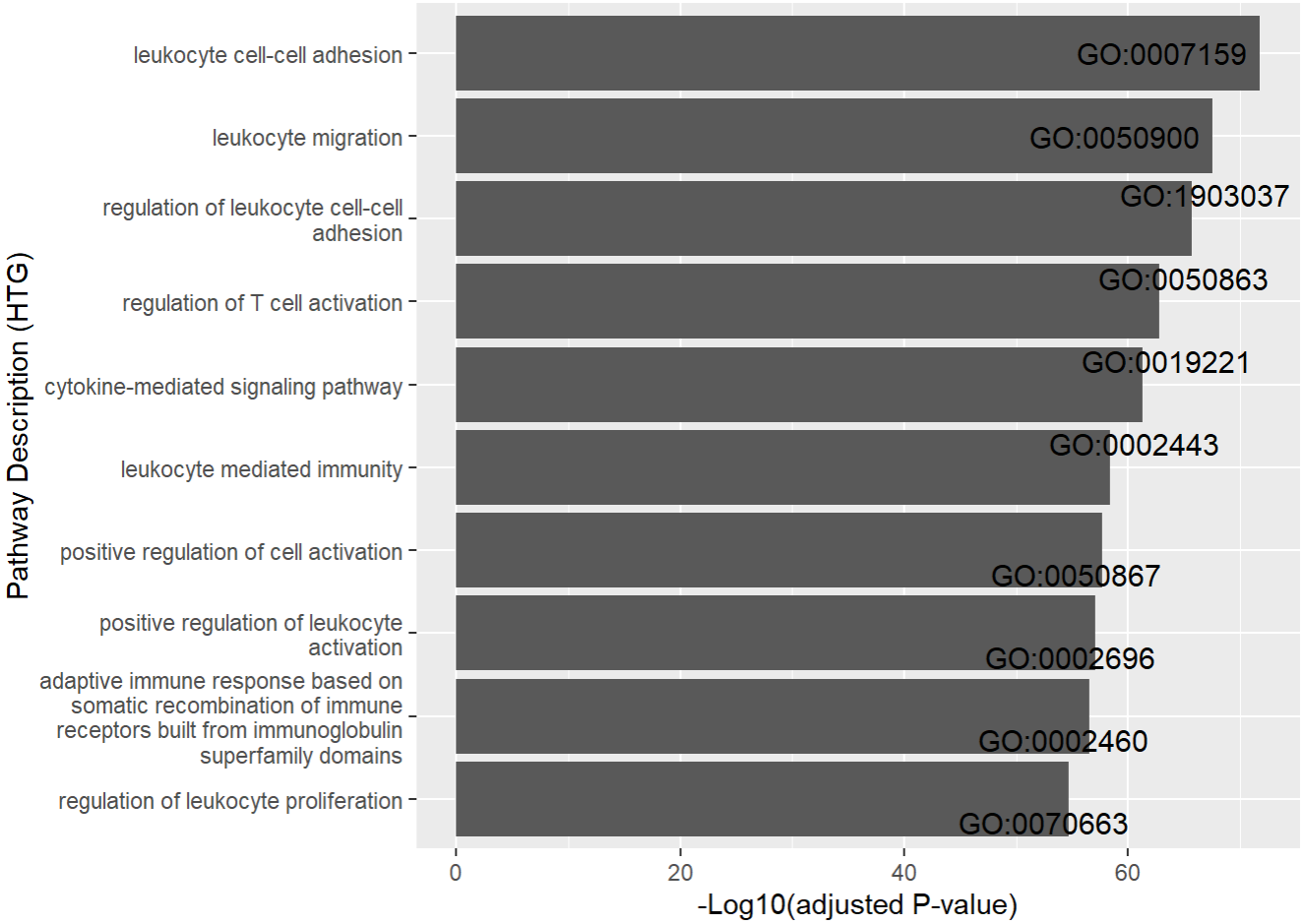
```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                              "Description"
## [3] "GeneRatio"                       "BgRatio"
## [5] "pvalue"                          "p.adjust"
## [7] "qvalue"                          "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_log10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
```

```
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTG)")
```




```
df <- results_goSEA_HTseq[1:10,]
head(df)
```

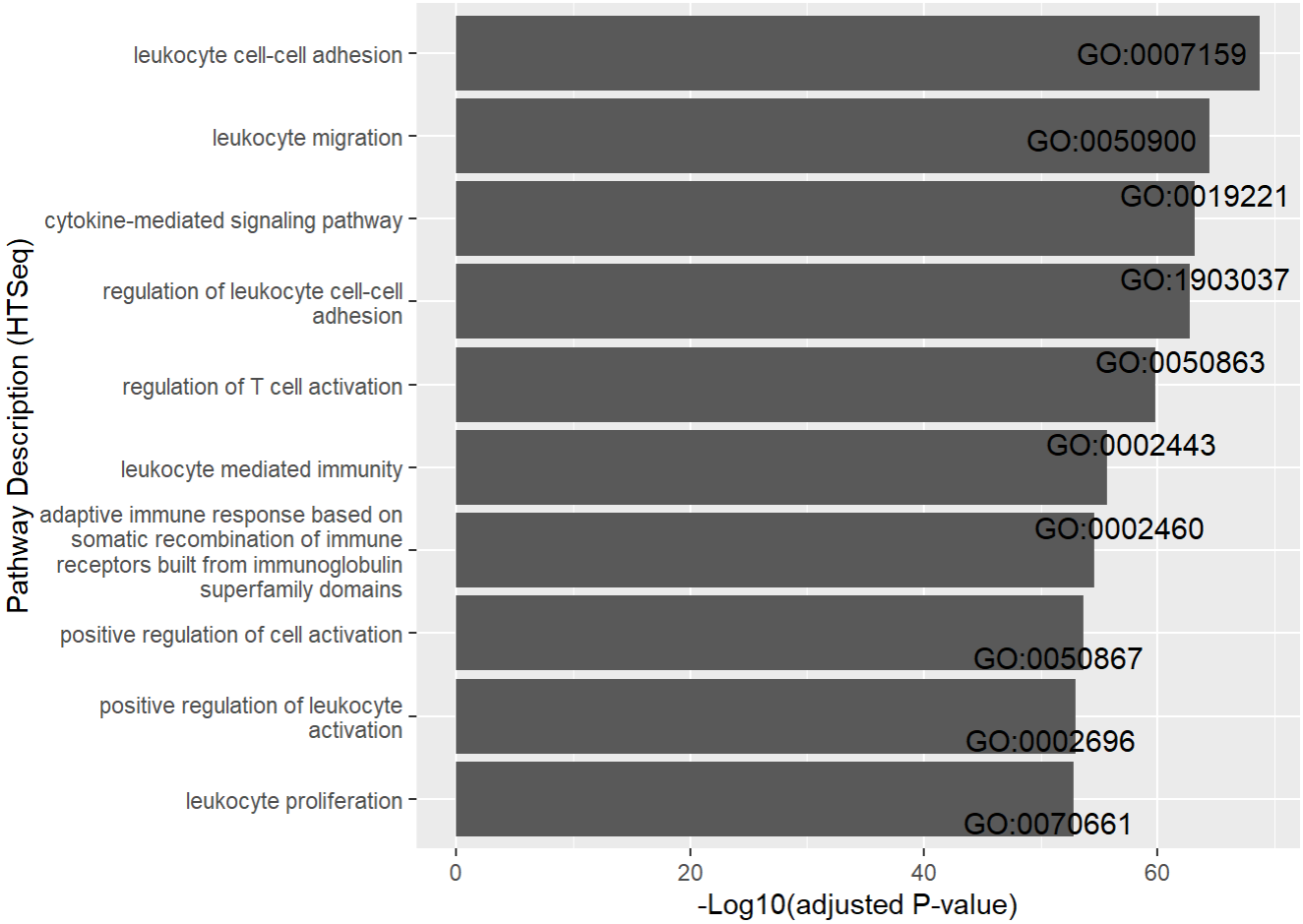
```
##                    ID                                Description GeneRatio
## GO:0007159 GO:0007159              leukocyte cell-cell adhesion    88/446
## GO:0050900 GO:0050900                       leukocyte migration    82/446
## GO:0019221 GO:0019221       cytokine-mediated signaling pathway    84/446
## GO:1903037 GO:1903037 regulation of leukocyte cell-cell adhesion    80/446
## GO:0050863 GO:0050863           regulation of T cell activation    78/446
## GO:0002443 GO:0002443               leukocyte mediated immunity    81/446
##               BgRatio       pvalue     p.adjust       qvalue
## GO:0007159 420/28564 3.866349e-73 1.833036e-69 7.264666e-70
## GO:0050900 386/28564 1.491275e-68 3.535068e-65 1.401014e-65
## GO:0019221 429/28564 3.959092e-67 6.256684e-64 2.479642e-64
## GO:1903037 380/28564 1.634343e-66 1.937105e-63 7.677111e-64
## GO:0050863 384/28564 1.542082e-63 1.462202e-60 5.794982e-61
## GO:0002443 480/28564 2.783988e-59 2.199815e-56 8.718278e-57
```

```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                                  "Description"
## [3] "GeneRatio"                           "BgRatio"
## [5] "pvalue"                              "p.adjust"
## [7] "qvalue"                              "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTSeq)")
```



Step 3

Run GO-SEA

HTSeq

ColOnly vs Colnodysp

```
#GO SEA
goSEA_HTseq_2_Colnodysp <- enrichGO(
  gene = rownames(sigRes_HTSeq_2_Colnodysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)


goSEA_HTseq <- goSEA_HTseq_2_Colnodysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTseq)
paste("Number of pathways differentially regulated is", nrow(goSEA_HTseq))
```

```
## [1] "Number of pathways differentially regulated is 182"
```

```
# goSEA_HTseq[1:10,1:7]
results_goSEA_HTseq <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjus
t", "qvalue"))
results_goSEA_HTseq <- t(results_goSEA_HTseq)
colnames(results_goSEA_HTseq) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust
", "qvalue"); results_goSEA_HTseq <- results_goSEA_HTseq[-1,]
results_goSEA_HTseq <- goSEA_HTseq[1:10,1:7]
```

HTG ColOnly vs Colnodysp

```
#GO SEA
goSEA_HTG_2_Colnodysp <- enrichGO(
  gene = rownames(sigRes_HTG_2_Colnodysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)

goSEA_HTG <- goSEA_HTG_2_Colnodysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTG)
paste("ColOnly vs Colitis without dysp,, HTG - Number of pathways differentially regulated is"
, nrow(goSEA_HTG))
```

```
## [1] "ColOnly vs Colitis without dysp,, HTG - Number of pathways differentially regulated is
 195"
```

```
paste("ColOnly vs Colitis without dysp, HTSeq - Number of pathways differentially regulated is
", nrow(goSEA_HTseq))
```

```
## [1] "ColOnly vs Colitis without dysp, HTSeq - Number of pathways differentially regulated i
s 182"
```

```
# goSEA_HTG[1:10,1:7]
results_goSEA_HTG <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust"
, "qvalue"))
results_goSEA_HTG <- t(results_goSEA_HTG)
colnames(results_goSEA_HTG) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust",
 "qvalue"); results_goSEA_HTG <- results_goSEA_HTG[-1,]
results_goSEA_HTG <- goSEA_HTG[1:10,1:7]
```

Then, compare the top 10 differentially regulated pathways in the nocol vs colnodysp groups

```
results3 <- data.frame(top10pathways_HTG=results_goSEA_HTG[,2],top10pathways_HTSeq=results_goS
EA_HTseq[,2])
results3
```

```
##                        top10pathways_HTG                  top10pathways_HTSeq
## 1                    neutrophil migration                neutrophil chemotaxis
## 2             myeloid leukocyte migration               granulocyte chemotaxis
## 3                   granulocyte migration                 neutrophil migration
## 4                   neutrophil chemotaxis                 leukocyte chemotaxis
## 5                  granulocyte chemotaxis          myeloid leukocyte migration
## 6                    leukocyte chemotaxis                granulocyte migration
## 7                     leukocyte migration             humoral immune response
## 8                 humoral immune response                        cell chemotaxis
## 9                        cell chemotaxis chemokine-mediated signaling pathway
## 10 chemokine-mediated signaling pathway                   leukocyte migration
```

```
write.csv(results3,file = "C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/
results/pipeline_comparisons/top10_diff_reg_pathways_nocol_coldysp.csv")
setdiff(results_goSEA_HTG[,2],results_goSEA_HTseq[,2])
```

```
## character(0)
```

```
setdiff(results_goSEA_HTseq[,2],results_goSEA_HTG[,2])
```

```
## character(0)
```

ggplot for visualisation

```
library(stringr)
# df <- goSEA_HTG@result[1:10,]
df <- results_goSEA_HTG[1:10,]
head(df)
```

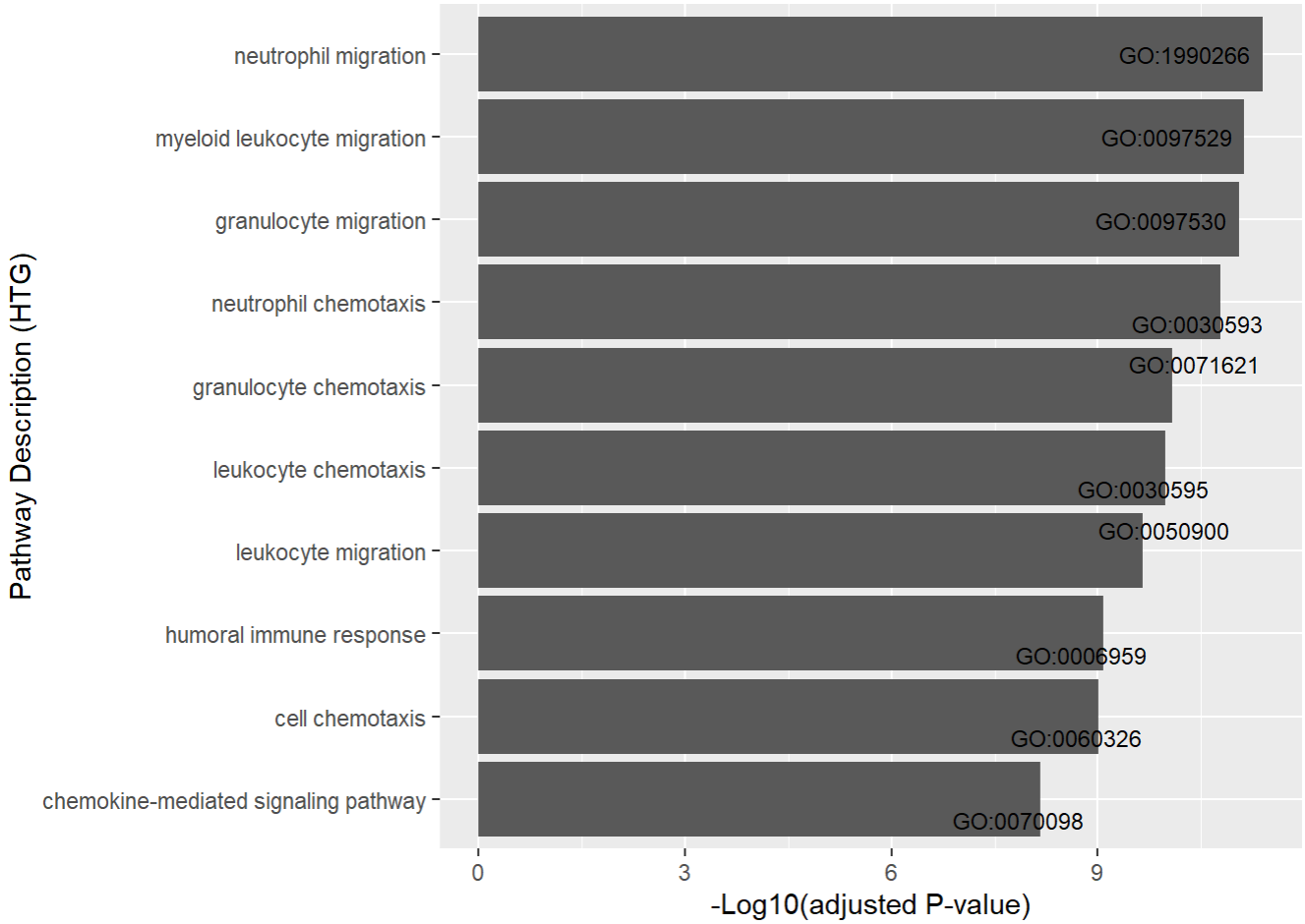
```
##                    ID                  Description GeneRatio   BgRatio
## GO:1990266 GO:1990266         neutrophil migration      9/27 132/28564
## GO:0097529 GO:0097529 myeloid leukocyte migration     10/27 242/28564
## GO:0097530 GO:0097530        granulocyte migration      9/27 162/28564
## GO:0030593 GO:0030593        neutrophil chemotaxis      8/27 105/28564
## GO:0071621 GO:0071621       granulocyte chemotaxis      8/27 131/28564
## GO:0030595 GO:0030595         leukocyte chemotaxis      9/27 230/28564
##                   pvalue     p.adjust       qvalue
## GO:1990266 3.182634e-15 3.984658e-12 2.030185e-12
## GO:0097529 1.175551e-14 7.358948e-12 3.749388e-12
## GO:0097530 2.082878e-14 8.692546e-12 4.428857e-12
## GO:0030593 5.322709e-14 1.666008e-11 8.488320e-12
## GO:0071621 3.251715e-13 8.142294e-11 4.148504e-11
## GO:0030595 5.025187e-13 1.048589e-10 5.342567e-11
```

```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                              "Description"
## [3] "GeneRatio"                       "BgRatio"
## [5] "pvalue"                          "p.adjust"
## [7] "qvalue"                          "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=3,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTG)")
```




```
df <- results_goSEA_HTseq[1:10,]
head(df)
```

```
##                    ID                 Description GeneRatio   BgRatio
## GO:0030593 GO:0030593       neutrophil chemotaxis      8/29 105/28564
## GO:0071621 GO:0071621     granulocyte chemotaxis      8/29 131/28564
## GO:1990266 GO:1990266       neutrophil migration      8/29 132/28564
## GO:0030595 GO:0030595       leukocyte chemotaxis      9/29 230/28564
## GO:0097529 GO:0097529 myeloid leukocyte migration      9/29 242/28564
## GO:0097530 GO:0097530      granulocyte migration      8/29 162/28564
##                   pvalue     p.adjust       qvalue
## GO:0030593 1.022853e-13 1.264246e-10 6.998468e-11
## GO:0071621 6.238617e-13 2.734343e-10 1.513646e-10
## GO:1990266 6.636755e-13 2.734343e-10 1.513646e-10
## GO:0030595 1.058907e-12 3.272022e-10 1.811288e-10
## GO:0097529 1.674150e-12 4.138499e-10 2.290942e-10
## GO:0097530 3.488247e-12 7.185788e-10 3.977825e-10
```

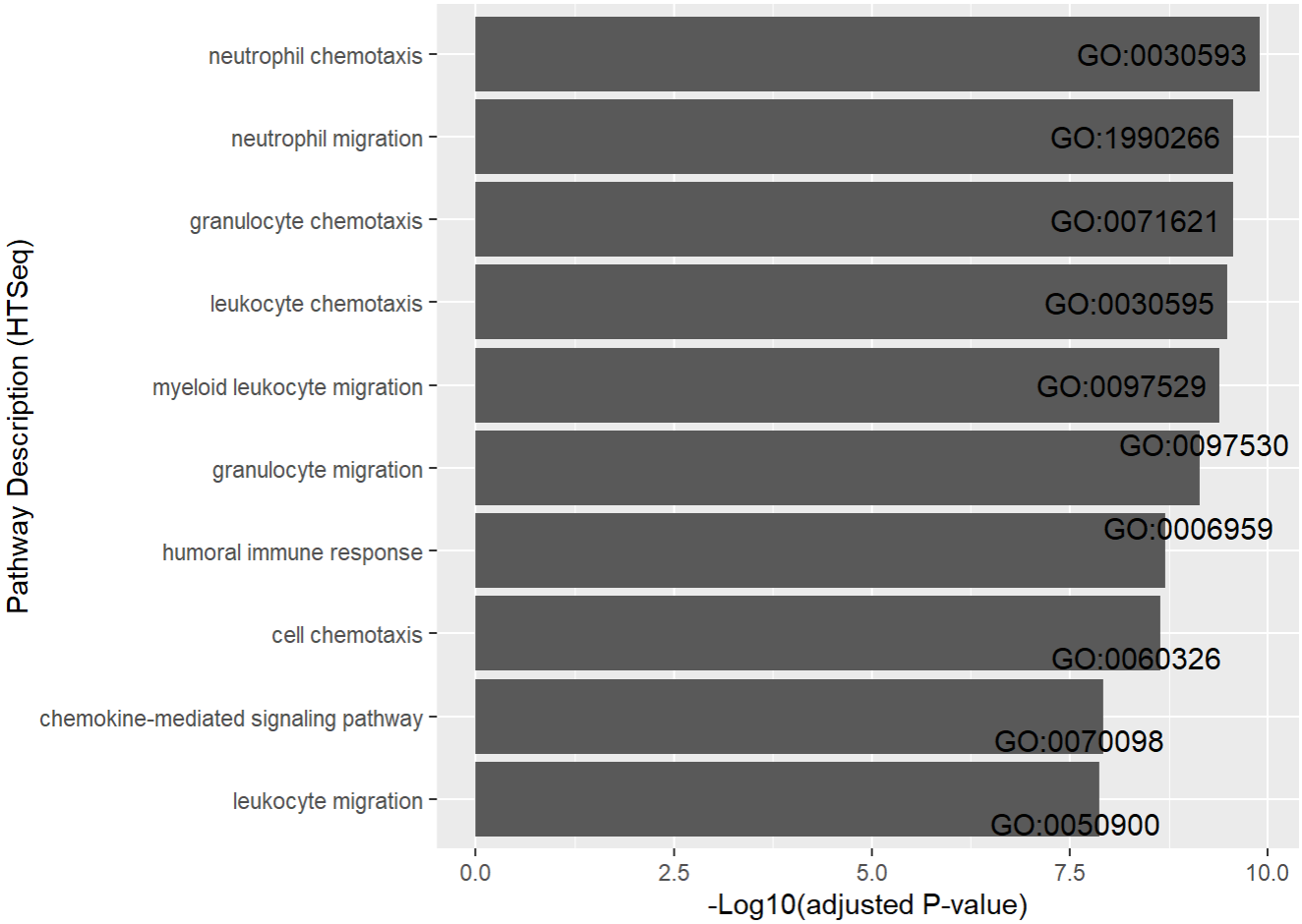
```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                          "Description"
## [3] "GeneRatio"                   "BgRatio"
```

```
## [5] "pvalue"                          "p.adjust"
## [7] "qvalue"                          "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTSeq)")
```



Step 3

Run GO-SEA

HTSeq

ColOnly vs Coldysp

```
#GO SEA
goSEA_HTseq_2_Coldysp <- enrichGO(
  gene = rownames(sigRes_HTSeq_2_Coldysp),
```

```
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)


goSEA_HTseq <- goSEA_HTseq_2_Coldysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTseq)
paste("Number of pathways differentially regulated is", nrow(goSEA_HTseq))
```

```
## [1] "Number of pathways differentially regulated is 122"
```

```
# goSEA_HTseq[1:10,1:7]
results_goSEA_HTseq <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjus
t", "qvalue"))
results_goSEA_HTseq <- t(results_goSEA_HTseq)
colnames(results_goSEA_HTseq) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust
", "qvalue"); results_goSEA_HTseq <- results_goSEA_HTseq[-1,]
results_goSEA_HTseq <- goSEA_HTseq[1:10,1:7]
```

HTG ColOnly vs Coldysp

```
#GO SEA
goSEA_HTG_2_Coldysp <- enrichGO(
  gene = rownames(sigRes_HTG_2_Coldysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)

goSEA_HTG <- goSEA_HTG_2_Coldysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTG)
paste("ColOnly vs Colitis with dysp,, HTG - Number of pathways differentially regulated is", n
row(goSEA_HTG))
```

```
## [1] "ColOnly vs Colitis with dysp,, HTG - Number of pathways differentially regulated is 54
```

```
"
```

```
paste("ColOnly vs Colitis with dysp, HTSeq - Number of pathways differentially regulated is",
nrow(goSEA_HTseq))
```

```
## [1] "ColOnly vs Colitis with dysp, HTSeq - Number of pathways differentially regulated is 1
22"
```

```
# goSEA_HTG[1:10,1:7]
results_goSEA_HTG <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust"
, "qvalue"))
results_goSEA_HTG <- t(results_goSEA_HTG)
colnames(results_goSEA_HTG) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust",
 "qvalue"); results_goSEA_HTG <- results_goSEA_HTG[-1,]
results_goSEA_HTG <- goSEA_HTG[1:10,1:7]
```

Then, compare the top 10 differentially regulated pathways in the nocol vs colnodysp groups

```
results3 <- data.frame(top10pathways_HTG=results_goSEA_HTG[,2],top10pathways_HTSeq=results_goS
EA_HTseq[,2])
results3
```

```
##                                           top10pathways_HTG
## 1                          cytokine-mediated signaling pathway
## 2                        chemokine-mediated signaling pathway
## 3                                     response to chemokine
## 4                           cellular response to chemokine
## 5                                     neutrophil chemotaxis
## 6                                   granulocyte chemotaxis
## 7                                     neutrophil migration
## 8                      positive regulation of defense response
## 9                                     granulocyte migration
## 10 positive regulation of cytosolic calcium ion concentration
##                                        top10pathways_HTSeq
## 1                       cytokine-mediated signaling pathway
## 2                                    humoral immune response
## 3                  positive regulation of defense response
## 4                     chemokine-mediated signaling pathway
## 5                                     response to chemokine
## 6                           cellular response to chemokine
## 7                                     neutrophil chemotaxis
## 8                           peptidyl-serine phosphorylation
## 9                             peptidyl-serine modification
## 10 positive regulation of peptidyl-serine phosphorylation
```

```
write.csv(results3,file = "C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/
results/pipeline_comparisons/top10_diff_reg_pathways_nocol_coldysp.csv")
setdiff(results_goSEA_HTG[,2],results_goSEA_HTseq[,2])
```

```
## [1] "granulocyte chemotaxis"
## [2] "neutrophil migration"
## [3] "granulocyte migration"
## [4] "positive regulation of cytosolic calcium ion concentration"
```

```
setdiff(results_goSEA_HTseq[,2],results_goSEA_HTG[,2])
```

```
## [1] "humoral immune response"
## [2] "peptidyl-serine phosphorylation"
## [3] "peptidyl-serine modification"
## [4] "positive regulation of peptidyl-serine phosphorylation"
```

ggplot for visualisation

```
library(stringr)
# df <- goSEA_HTG@result[1:10,]
df <- results_goSEA_HTG[1:10,]
head(df)
```

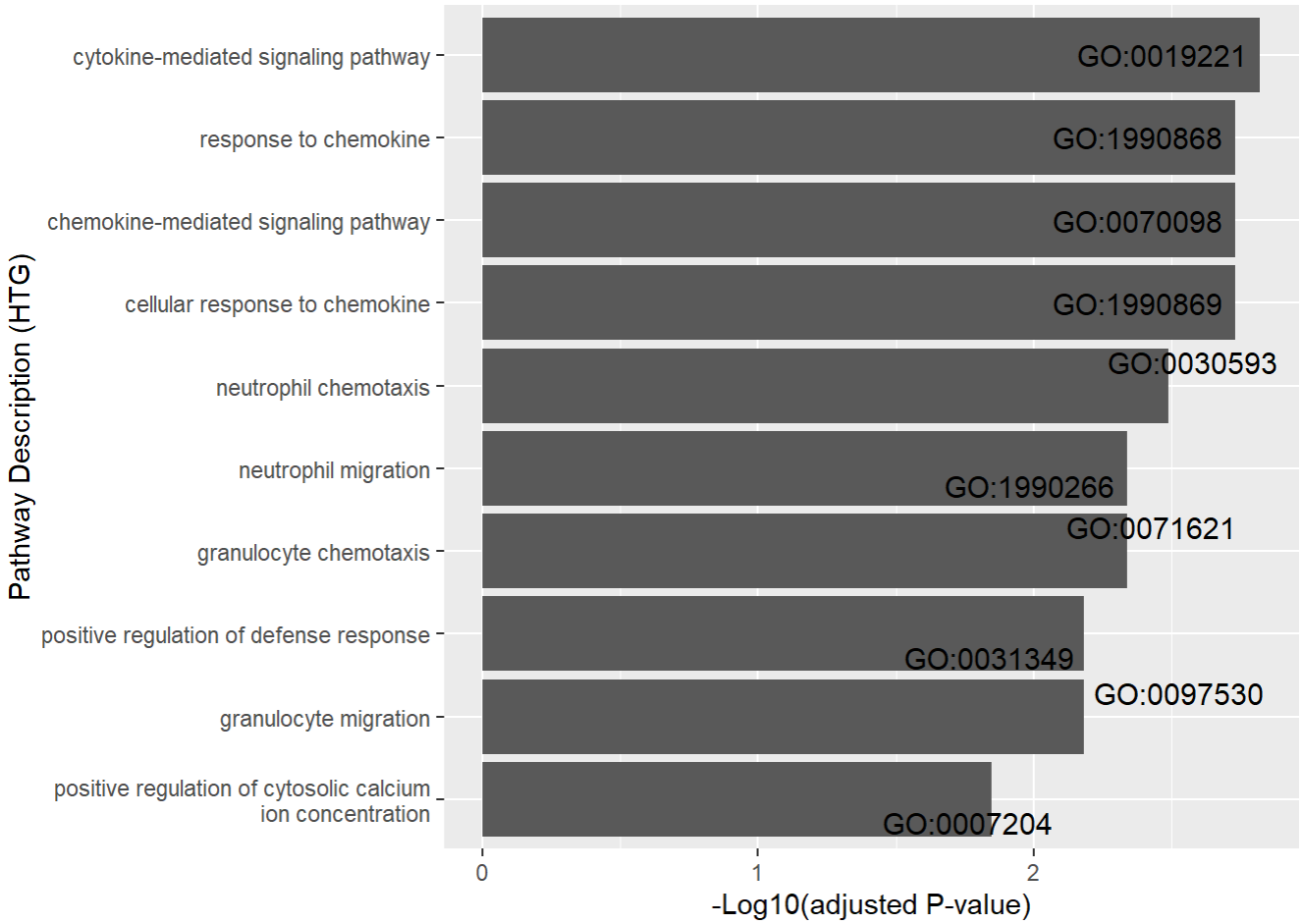
```
##                    ID                          Description GeneRatio   BgRatio
## GO:0019221 GO:0019221  cytokine-mediated signaling pathway      5/15 429/28564
## GO:0070098 GO:0070098 chemokine-mediated signaling pathway      3/15  71/28564
## GO:1990868 GO:1990868               response to chemokine      3/15  81/28564
## GO:1990869 GO:1990869      cellular response to chemokine      3/15  81/28564
## GO:0030593 GO:0030593               neutrophil chemotaxis      3/15 105/28564
## GO:0071621 GO:0071621              granulocyte chemotaxis      3/15 131/28564
##                  pvalue    p.adjust       qvalue
## GO:0019221 1.980407e-06 0.001503129 0.0007629777
## GO:0070098 6.553840e-06 0.001850567 0.0009393350
## GO:1990868 9.752658e-06 0.001850567 0.0009393350
## GO:1990869 9.752658e-06 0.001850567 0.0009393350
## GO:0030593 2.126639e-05 0.003228238 0.0016386313
## GO:0071621 4.119806e-05 0.004569490 0.0023194414
```

```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                          "Description"
## [3] "GeneRatio"                   "BgRatio"
## [5] "pvalue"                      "p.adjust"
## [7] "qvalue"                      "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +
```

```
  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTG)")
```




```
df <- results_goSEA_HTseq[1:10,]
head(df)
```

```
##                    ID                          Description GeneRatio
## GO:0019221 GO:0019221     cytokine-mediated signaling pathway      7/17
## GO:0006959 GO:0006959                 humoral immune response      5/17
## GO:0031349 GO:0031349 positive regulation of defense response      5/17
## GO:0070098 GO:0070098    chemokine-mediated signaling pathway      3/17
## GO:1990868 GO:1990868                   response to chemokine      3/17
## GO:1990869 GO:1990869          cellular response to chemokine      3/17
##               BgRatio       pvalue     p.adjust       qvalue
## GO:0019221 429/28564 2.804263e-09 2.338755e-06 1.133512e-06
## GO:0006959 299/28564 6.784501e-07 2.829137e-04 1.371183e-04
## GO:0031349 452/28564 5.132985e-06 1.426970e-03 6.916021e-04
## GO:0070098  71/28564 9.759840e-06 2.017701e-03 9.779089e-04
## GO:1990868  81/28564 1.451584e-05 2.017701e-03 9.779089e-04
## GO:1990869  81/28564 1.451584e-05 2.017701e-03 9.779089e-04
```

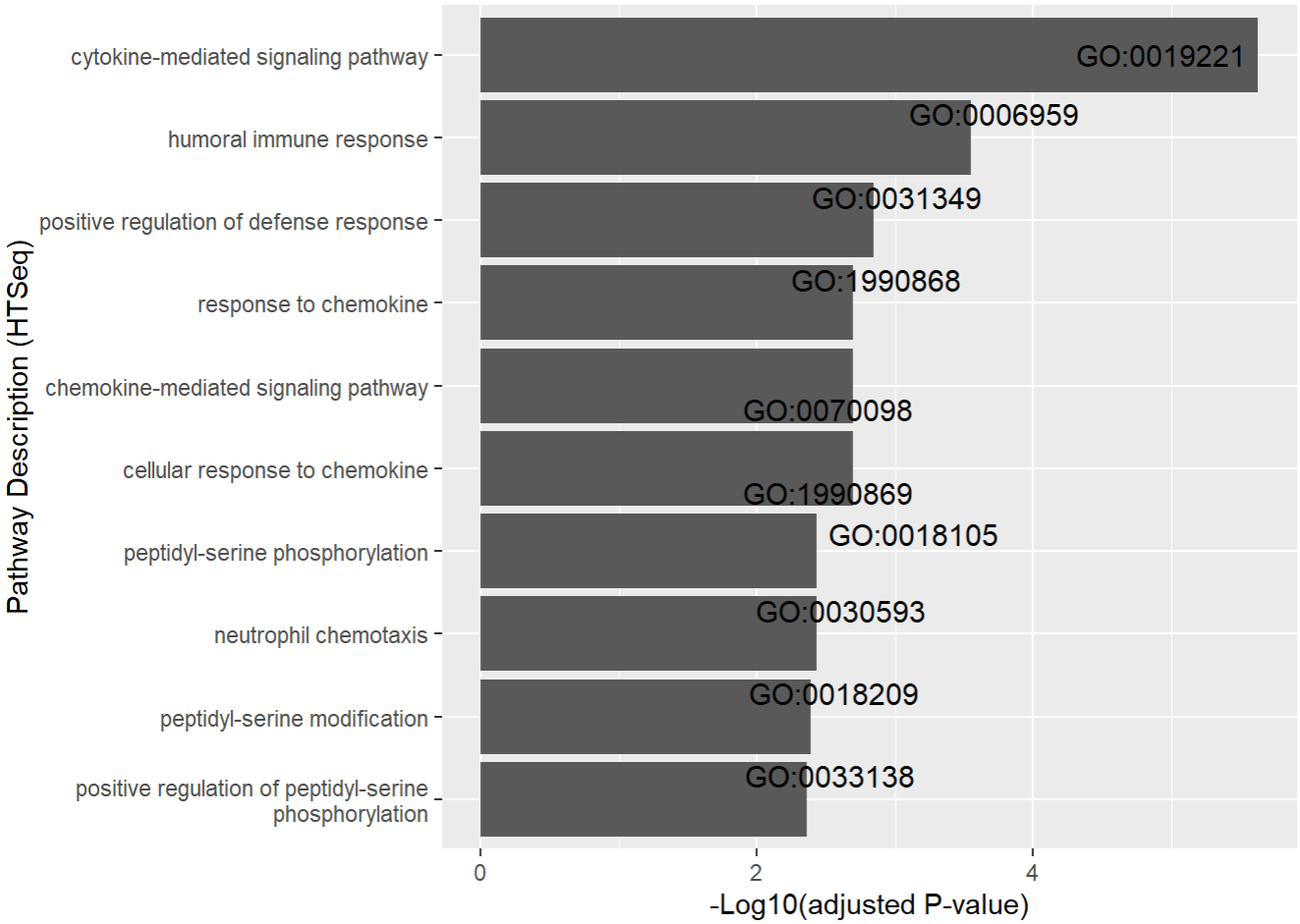
```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
```

```
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                                "Description"
## [3] "GeneRatio"                         "BgRatio"
## [5] "pvalue"                            "p.adjust"
## [7] "qvalue"                            "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTSeq)")
```



Step 3

Run GO-SEA

HTSeq Colnodysp vs Coldysp

```
#GO SEA
goSEA_HTseq_3_Colnodysp <- enrichGO(
  gene = rownames(sigRes_HTSeq_3_Colnodysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)


goSEA_HTseq <- goSEA_HTseq_3_Colnodysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTseq)
paste("Number of pathways differentially regulated is", nrow(goSEA_HTseq))
```

```
## [1] "Number of pathways differentially regulated is 794"
```

```
# goSEA_HTseq[1:10,1:7]
results_goSEA_HTseq <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjus
t", "qvalue"))
results_goSEA_HTseq <- t(results_goSEA_HTseq)
colnames(results_goSEA_HTseq) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust
", "qvalue"); results_goSEA_HTseq <- results_goSEA_HTseq[-1,]
results_goSEA_HTseq <- goSEA_HTseq[1:10,1:7]
```

HTG WT vs Coldysp

```
#GO SEA
goSEA_HTG_3_Coldysp <- enrichGO(
  gene = rownames(sigRes_HTG_3_Colnodysp),
  OrgDb = org.Mm.eg.db,
  keyType = "SYMBOL",
  ont = "BP", #BP, MF or CC
  pAdjustMethod = "BH",
  pvalueCutoff = 0.05,
  qvalueCutoff = 0.05)

goSEA_HTG <- goSEA_HTG_3_Coldysp
```

Save the top 10 differentially regulated pathways

```
# class(goSEA_HTG)
# dim(goSEA_HTG)
paste("Col without dysp vs Colitis with dysp, HTG - Number of pathways differentially regulate
d is", nrow(goSEA_HTG))
```

```
## [1] "Col without dysp vs Colitis with dysp, HTG - Number of pathways differentially regulat
ed is 826"
```

```
paste("Col without dysp vs Colitis with dysp, HTSeq - Number of pathways differentially regula
ted is", nrow(goSEA_HTseq))
```

```
## [1] "Col without dysp vs Colitis with dysp, HTSeq - Number of pathways differentially regul
ated is 794"
```

```
# goSEA_HTG[1:10,1:7]
results_goSEA_HTG <- data.frame(c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust"
, "qvalue"))
results_goSEA_HTG <- t(results_goSEA_HTG)
colnames(results_goSEA_HTG) <- c("ID","Description","GeneRatio","BgRatio","pvalue","p.adjust",
 "qvalue"); results_goSEA_HTG <- results_goSEA_HTG[-1,]
results_goSEA_HTG <- goSEA_HTG[1:10,1:7]
```

Then, compare the top 10 differentially regulated pathways in the nocol vs colnodysp groups

```
results3 <- data.frame(top10pathways_HTG=results_goSEA_HTG[,2],top10pathways_HTSeq=results_goS
EA_HTseq[,2])
results3
```

```
##                       top10pathways_HTG                  top10pathways_HTSeq
## 1           myeloid leukocyte migration         myeloid leukocyte migration
## 2                  leukocyte chemotaxis                leukocyte chemotaxis
## 3                   leukocyte migration               granulocyte migration
## 4                 granulocyte migration                 leukocyte migration
## 5                       cell chemotaxis                     cell chemotaxis
## 6                granulocyte chemotaxis cytokine-mediated signaling pathway
## 7                  neutrophil migration              granulocyte chemotaxis
## 8                 neutrophil chemotaxis                neutrophil migration
## 9  cytokine-mediated signaling pathway               neutrophil chemotaxis
## 10              humoral immune response             humoral immune response
```

```
write.csv(results3,file = "C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/
results/pipeline_comparisons/top10_diff_reg_pathways_nocol_coldysp.csv")
setdiff(results_goSEA_HTG[,2],results_goSEA_HTseq[,2])
```

```
## character(0)
```

```
setdiff(results_goSEA_HTseq[,2],results_goSEA_HTG[,2])
```

```
## character(0)
```

ggplot for visualisation

```
library(stringr)
# df <- goSEA_HTG@result[1:10,]
df <- results_goSEA_HTG[1:10,]
head(df)
```

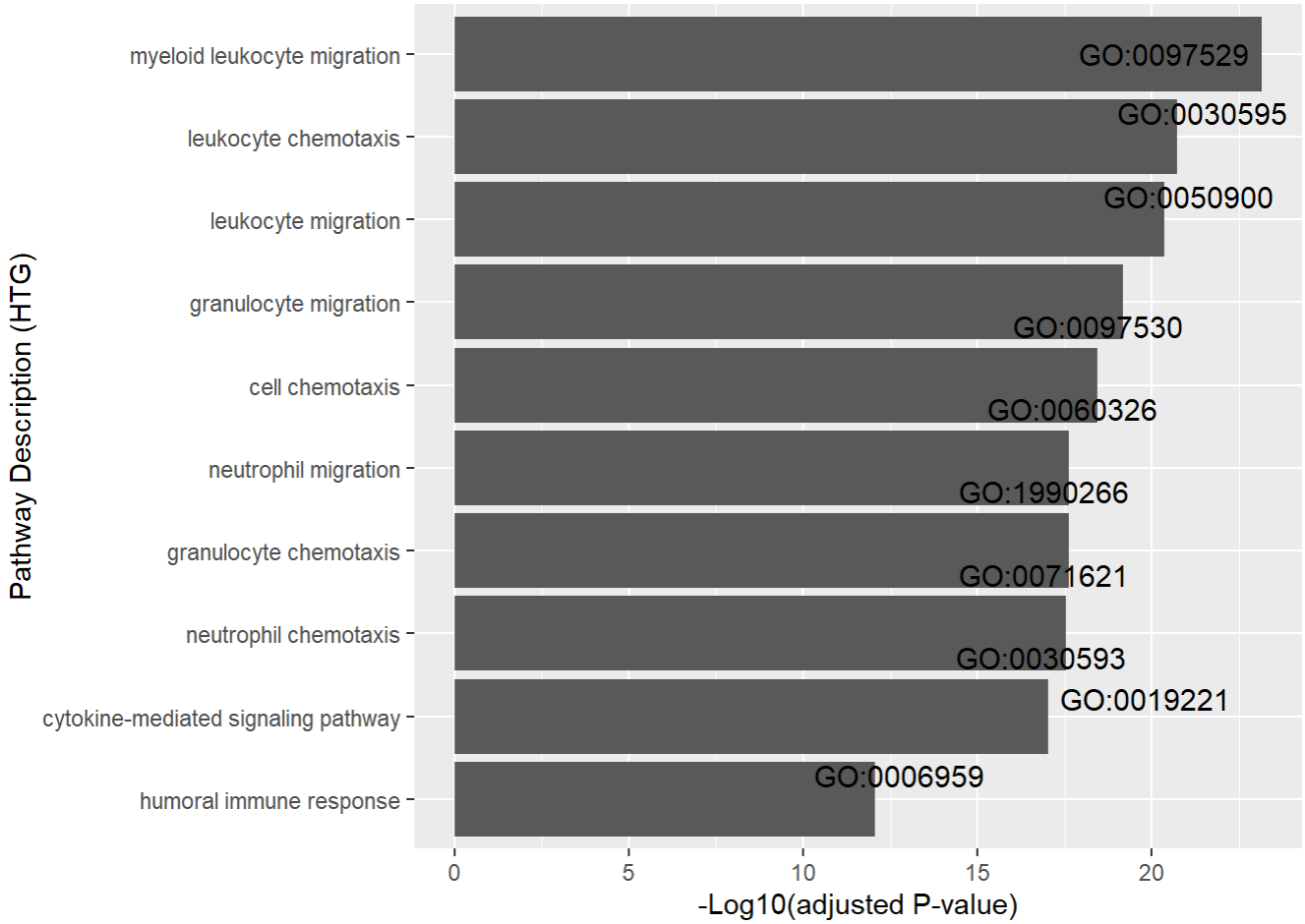
```
##                    ID                  Description GeneRatio   BgRatio
## GO:0097529 GO:0097529 myeloid leukocyte migration     23/95 242/28564
## GO:0030595 GO:0030595        leukocyte chemotaxis     21/95 230/28564
## GO:0050900 GO:0050900         leukocyte migration     24/95 386/28564
## GO:0097530 GO:0097530       granulocyte migration     18/95 162/28564
## GO:0060326 GO:0060326             cell chemotaxis     21/95 308/28564
## GO:0071621 GO:0071621      granulocyte chemotaxis     16/95 131/28564
##                   pvalue     p.adjust       qvalue
## GO:0097529 2.901470e-27 7.079586e-24 3.936836e-24
## GO:0030595 1.510253e-24 1.842509e-21 1.024587e-21
## GO:0050900 5.487643e-24 4.463283e-21 2.481955e-21
## GO:0097530 1.068049e-22 6.515102e-20 3.622936e-20
## GO:0060326 7.307389e-22 3.566006e-19 1.982995e-19
## GO:0071621 6.052870e-21 2.394602e-18 1.331597e-18
```

```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                                "Description"
## [3] "GeneRatio"                         "BgRatio"
## [5] "pvalue"                            "p.adjust"
## [7] "qvalue"                            "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTG)")
```




```
df <- results_goSEA_HTseq[1:10,]
head(df)
```

```
##                    ID                         Description GeneRatio   BgRatio
## GO:0097529 GO:0097529         myeloid leukocyte migration     22/92 242/28564
## GO:0030595 GO:0030595               leukocyte chemotaxis     21/92 230/28564
## GO:0097530 GO:0097530              granulocyte migration     18/92 162/28564
## GO:0050900 GO:0050900                leukocyte migration     23/92 386/28564
## GO:0060326 GO:0060326                    cell chemotaxis     21/92 308/28564
## GO:0019221 GO:0019221 cytokine-mediated signaling pathway     23/92 429/28564
##                  pvalue     p.adjust       qvalue
## GO:0097529 5.401331e-26 1.321706e-22 7.487950e-23
## GO:0030595 7.223589e-25 8.838062e-22 5.007088e-22
## GO:0097530 5.726288e-23 3.882086e-20 2.199345e-20
## GO:0050900 6.345870e-23 3.882086e-20 2.199345e-20
## GO:0060326 3.522832e-22 1.724074e-19 9.767516e-20
## GO:0019221 6.971400e-22 2.843169e-19 1.610760e-19
```

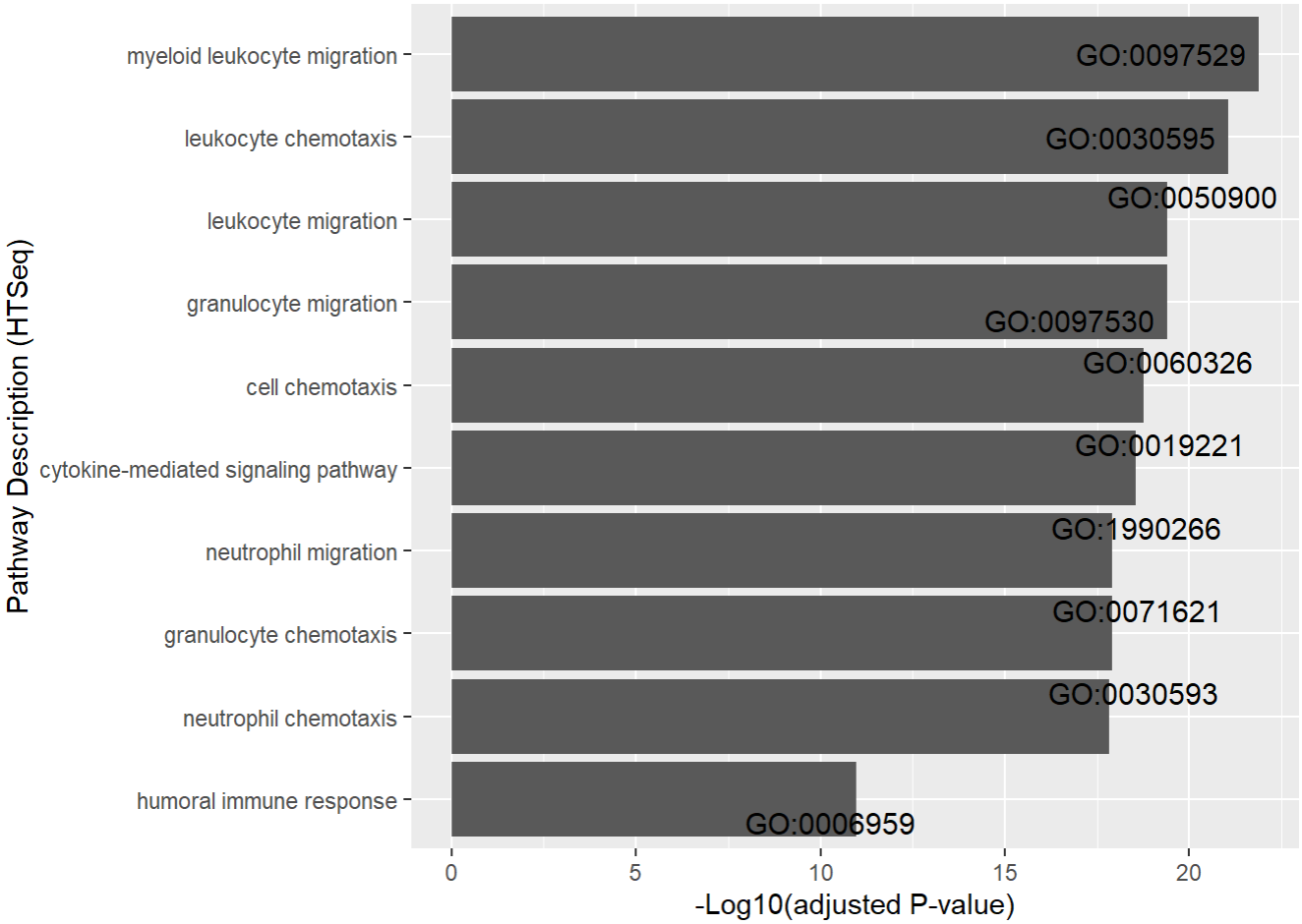
```
df$Negative_log10_adjusted_PValue <- -log10(df$p.adjust)
df$Pathway_Description <- str_wrap(df$Description, width = 40)
colnames(df)
```

```
## [1] "ID"                          "Description"
## [3] "GeneRatio"                   "BgRatio"
```

```
## [5] "pvalue"                          "p.adjust"
## [7] "qvalue"                          "Negative_log10_adjusted_PValue"
## [9] "Pathway_Description"
```

```
ggplot(df, aes(x = Negative_log10_adjusted_PValue, y = reorder(Pathway_Description, Negative_l
og10_adjusted_PValue), label=ID)) +

  geom_bar(stat = "identity")+
  geom_text_repel(
    colour="black",size=4,min.segment.length = 10,
    hjust = 0.5, vjust = 0.5)+
  labs(x = "-Log10(adjusted P-value)", y = "Pathway Description (HTSeq)")
```



Save image

```
save.image("C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.23/scripts/DEG_ana
lysis/Gene_counts/GSEA/GSEA.RData")
```

Attmepting to convert the gene symbols to ENTEREZID first

```
# #Convert symbols to Entrez IDs using bitr (Biological ID translator)
# sigRes$Entrez <- bitr(
#   rownames(sigRes),
#   fromType = "SYMBOL",
```

```
#   toType = c("ENTREZID"),
#   OrgDb = org.Mm.eg.db,
#   drop = FALSE)$ENTREZID


gene_names <- data.frame(column1=c("Bcl2a1a_Bcl2a1d","Bcl2a1b","Ccl21_family","Ccl21a","Ccl27_
family","Ccl27a","Cd244","Cd244a"))
gene_names <- c("Bcl2a1a_Bcl2a1d","Bcl2a1b","Ccl21_family","Ccl21a","Ccl27_family","Ccl27a","C
d244","Cd244a")

gene_names <- c("Bcl2a1a_Bcl2a1d","Bcl2a1d","Bcl2a1b","Ccl21_family","Ccl21a","Ccl27_family","
Ccl27a","Cd244","Cd244a","H2-Ea-ps","H2-Ea","Ifit3_Ifit3b","Ifit3","Ifna_family","Ifna1","Ifna
15","Ifna6","Ifnab","Ifnl2_Ifnl3","Ifnl2","Ifnl3","Il22_Iltifb","Il22","Il22b","Klra20","Klra2
0","Klra21","Klra4_Klra18","Klra7_Klra20","Klra13-ps","Klra4","Klra7","Lyz1_Lyz2","Lyz1","Lyz2
","Oas1a_Oas1g","Oas1a","Oas1g","Prame","Pramex1","Skp1a","Skp1")

gene_names$Entrez <- bitr(
  gene_names,
  fromType = "SYMBOL",
  toType = c("ENTREZID"),
  OrgDb = org.Mm.eg.db,
  drop = FALSE)#$ENTREZID
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Warning in bitr(gene_names, fromType = "SYMBOL", toType = c("ENTREZID"), :
## 36.59% of input gene IDs are fail to map...
```

```
## Warning in gene_names$Entrez <- bitr(gene_names, fromType = "SYMBOL", toType =
## c("ENTREZID"), : Coercing LHS to a list
```

```
gene_names2 <- c("BCL2A1A_BCL2A1D","BCL2A1D","BCL2A1B","CCL21_FAMILY","CCL21A","CCL27_FAMILY",
"CCL27A","CD244","CD244A","H2-EA-PS","H2-EA","IFIT3_IFIT3B","IFIT3","IFNA_FAMILY","IFNA1","IFN
A15","IFNA6","IFNAB","IFNL2_IFNL3","IFNL2","IFNL3","IL22_ILTIFB","IL22","IL22B","KLRA20","KLRA
20","KLRA21","KLRA4_KLRA18","KLRA7_KLRA20","KLRA13-PS","KLRA4","KLRA7","LYZ1_LYZ2","LYZ1","LYZ
2","OAS1A_OAS1G","OAS1A","OAS1G","PRAME","PRAMEX1","SKP1A","SKP1")

gene_names2$Entrez <- bitr(
  gene_names2,
  fromType = "SYMBOL",
  toType = c("ENTREZID"),
  OrgDb = org.Hs.eg.db,
  drop = FALSE)
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Warning in bitr(gene_names2, fromType = "SYMBOL", toType = c("ENTREZID"), :
## 78.05% of input gene IDs are fail to map...
```

```
## Warning in gene_names2$Entrez <- bitr(gene_names2, fromType = "SYMBOL", :
## Coercing LHS to a list
```

```
comparison <- data.frame(gene_names=gene_names$Entrez[,1], EntrezID_Mm=gene_names$Entrez[,2],E
ntrezID_Hs=gene_names2$Entrez[,2]); comparison
```

```
##         gene_names EntrezID_Mm EntrezID_Hs
## 1  Bcl2a1a_Bcl2a1d        <NA>        <NA>
## 2          Bcl2a1d       12047        <NA>
## 3          Bcl2a1b       12045        <NA>
## 4     Ccl21_family        <NA>        <NA>
## 5           Ccl21a       18829        <NA>
## 6     Ccl27_family        <NA>        <NA>
## 7           Ccl27a       20301        <NA>
## 8            Cd244        <NA>       51744
## 9           Cd244a       18106        <NA>
## 10        H2-Ea-ps        <NA>        <NA>
## 11           H2-Ea   100504404        <NA>
## 12    Ifit3_Ifit3b        <NA>        <NA>
## 13           Ifit3       15959        3437
## 14     Ifna_family        <NA>        <NA>
## 15           Ifna1       15962        3439
## 16          Ifna15      242517        <NA>
## 17           Ifna6       15969        3443
## 18           Ifnab       15974        <NA>
## 19     Ifnl2_Ifnl3        <NA>        <NA>
## 20           Ifnl2      330496      282616
## 21           Ifnl3      338374      282617
## 22     Il22_Iltifb        <NA>        <NA>
## 23            Il22       50929       50616
## 24           Il22b      116849        <NA>
## 25          Klra20       93967        <NA>
## 26          Klra21       93968        <NA>
## 27    Klra4_Klra18        <NA>        <NA>
## 28    Klra7_Klra20        <NA>        <NA>
## 29       Klra13-ps       16631        <NA>
## 30           Klra4       16635        <NA>
## 31           Klra7       16638        <NA>
## 32       Lyz1_Lyz2        <NA>        <NA>
## 33            Lyz1       17110        <NA>
## 34            Lyz2       17105        <NA>
## 35     Oas1a_Oas1g        <NA>        <NA>
## 36           Oas1a      246730        <NA>
## 37           Oas1g       23960        <NA>
## 38           Prame        <NA>       23532
## 39         Pramex1       75829        <NA>
## 40           Skp1a        <NA>        <NA>
## 41            Skp1       21402        6500
```

```
write.csv(comparison,file = "C:/hne_files/MSc_Bioinf/MScProject-main_C/MScProject-main_10.06.2
3/results/pipeline_comparisons/entrezID.csv")
```

```
x <- org.Hs.egSYMBOL2EG
# Get the entrez gene identifiers that are mapped to a gene symbol
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the entrez gene ID for the first five genes
  xx[1:5]
  # Get the first one
  xx[[1]]
}
```

```
## [1] "1"
```

```
## then run enrichGO once more
#
# goSEA_HTseq <- enrichGO(
#   gene = rownames(sigRes_HTSeq),
#   OrgDb = org.Mm.eg.db,
#   keyType = "SYMBOL",
#   ont = "BP", #BP, MF or CC
#   pAdjustMethod = "BH",
#   pvalueCutoff = 0.05,
#   qvalueCutoff = 0.05)




# #Convert symbols to Entrez IDs using bitr (Biological ID translator)

trial_sigRes_HTSeq <- sigRes_HTSeq
head(rownames(trial_sigRes_HTSeq))
```

```
## [1] "Il18bp"   "Ido1"     "Ltf"      "Nos2"     "Tnfrsf1b" "Gbp5"
```

```
trial_sigRes_HTSeq$Entrez <- bitr(
  rownames(trial_sigRes_HTSeq),
  fromType = "SYMBOL",
  toType = c("ENTREZID"),
  OrgDb = org.Mm.eg.db,
  drop = FALSE)
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Warning in bitr(rownames(trial_sigRes_HTSeq), fromType = "SYMBOL", toType =
## c("ENTREZID"), : 0.24% of input gene IDs are fail to map...
```

```
trial_sigRes_HTG <- sigRes_HTG
```

```
head(rownames(trial_sigRes_HTG))
```

```
## [1] "Il18bp"   "Ido1"     "Tnfrsf1b" "Ltf"      "Nos2"     "H2-Ab1"
```

```
trial_sigRes_HTG$Entrez <- bitr(
  rownames(trial_sigRes_HTG),
  fromType = "SYMBOL",
  toType = c("ENTREZID"),
  OrgDb = org.Mm.eg.db,
  drop = FALSE)
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
## Warning in bitr(rownames(trial_sigRes_HTG), fromType = "SYMBOL", toType =
## c("ENTREZID"), : 2.05% of input gene IDs are fail to map...
```

```
# trial_sigRes_HTSeq_entrez <- enrichGO(
#   gene = rownames(trial_sigRes_HTSeq),
#   OrgDb = org.Mm.eg.db,
#   keyType = "ENTREZID",
#   ont = "BP",
#   pvalueCutoff = 0.05,
#   pAdjustMethod = "BH",
#   qvalueCutoff = 0.05
#   )

# enrichGO(
#   gene = rownames(sigRes_HTSeq),
#   OrgDb = org.Mm.eg.db,
#   keyType = "ENTREZID",
#   ont = "BP",
#   pvalueCutoff = 0.05,
#   pAdjustMethod = "BH",
#   universe,
#   qvalueCutoff = 0.05,
#   minGSSize = 10,
#   maxGSSize = 500,
#   readable = FALSE,
#   pool = FALSE
# )
```