6-Gene Ontology enrichment

BAI Qiang*

2021-10-01 13:17:43 +0200

Contents

*University Liege, mail qiang.bai@uliege.be

1 Description

The DE gene lists used for enrichment analyses were calculated using Seurat function FindMarkers[1] with default arguments except only.pos = TRUE in order to output only positively regulated genes and logfc.threshold of 0.2. KEGG and Gene Ontology enrichment analyses were made using enrichKEGG and enrichGO functions from package clusterProfiler package[2] with default arguments.

2 Load data and packages

```
library(Seurat)
library(ggplot2)
library(dplyr)
results <- readRDS(file = "../3-Merge and cell typing/so.merged_clusters.
    seuratObject.Rds")
```

3 DE genes and GO enrichment

3.1 Cluster 1 DE analysis and GO enrichment

The Cluster 1 mentioned in manuscript refers to cluster 0 in the codes following. The same for cluster 2-4 in manuscript, they are cluster 1-3 in the codes.

Calculate the DE genes.

```
de_ident0 <- FindMarkers(results, ident.1 = "0", only.pos = TRUE, logfc.
    threshold = 0.2)
head(de_ident0)
```

```
## # A tibble: 6 x 5
##   p_val avg_log2FC pct.1 pct.2 p_val_adj
##   <dbl>      <dbl> <dbl> <dbl>     <dbl>
## 1     0      0.689 1     0.935         0
## 2     0      0.682 0.994 0.892         0
## 3     0      0.622 0.998 0.912         0
## 4     0      0.379 0.993 0.941         0
## 5     0      0.733 0.477 0.107         0
## 6     0      0.670 0.994 0.955         0
```

Save data

```
write.csv(de_ident0, file = "./Results_lists/DE_cluster1.csv")
```

KEGG/GO enrichment:

```
library(clusterProfiler)
source("../../R/entrez2symbol.R")
source("../../R/replaceEntrezID.R")

de_entrez.ident0 <- bitr( geneID = rownames(de_ident0), fromType = "SYMBOL",
    toType = "ENTREZID", OrgDb = "org.Hs.eg.db", drop = TRUE ) $ ENTREZID
result.enrichKEGG.de_ident0 <- enrichKEGG(de_entrez.ident0, organism = "hsa",
    keyType = "ncbi-geneid")
result.enrichKEGG.de_ident0 <- replaceEntrezID(result.enrichKEGG.de_ident0)
```

```
result.enrichKEGG.de_ident0@result
```

```
## # A tibble: 199 x 9
##    ID       Description  GeneRatio BgRatio  pvalue p.adjust  qvalue geneID
   Count
##    <chr>    <chr>        <chr>     <chr>     <dbl>    <dbl>   <dbl> <chr>
   <int>
##  1 hsa05150 Staphylococ~ 12/99     96/8095 1.52e-9  3.02e-7 2.61e-7 C1QA/~
      12
##  2 hsa04940 Type I diab~ 8/99      43/8095 3.83e-8  3.81e-6 3.29e-6 CPE/H~
       8
##  3 hsa05330 Allograft r~ 7/99      38/8095 3.06e-7  1.91e-5 1.65e-5 HLA-C~
       7
##  4 hsa04612 Antigen pro~ 9/99      78/8095 3.85e-7  1.91e-5 1.65e-5 CD74/~
       9
##  5 hsa05332 Graft-versu~ 7/99      42/8095 6.28e-7  2.50e-5 2.16e-5 HLA-C~
       7
##  6 hsa04610 Complement ~ 9/99      85/8095 8.10e-7  2.69e-5 2.32e-5 C1QA/~
       9
##  7 hsa05320 Autoimmune ~ 7/99      53/8095 3.22e-6  9.14e-5 7.88e-5 HLA-C~
       7
##  8 hsa05416 Viral myoca~ 7/99      60/8095 7.51e-6  1.87e-4 1.61e-4 HLA-C~
       7
##  9 hsa04514 Cell adhesi~ 10/99     149/80~ 1.27e-5  2.80e-4 2.41e-4 HLA-C~
      10
## 10 hsa04145 Phagosome    10/99     152/80~ 1.51e-5  3.00e-4 2.59e-4 MARCO~
      10
## # ... with 189 more rows
```

```
write.csv(result.enrichKEGG.de_ident0, file = "./Results_lists/enrichKEGG_DE_
   cluster1.csv")
```

```
result.enrichGO.de_ident0 <- enrichGO(gene = de_entrez.ident0, OrgDb = "org.Hs
   .eg.db", ont = "BP")
result.enrichGO.de_ident0 <- replaceEntrezID(result.enrichGO.de_ident0)
result.enrichGO.de_ident0@result
```

```
## # A tibble: 2,919 x 9
##    ID     Description GeneRatio BgRatio   pvalue p.adjust   qvalue geneID
   Count
##    <chr>  <chr>       <chr>     <chr>      <dbl>    <dbl>    <dbl> <chr>
   <int>
##  1 GO:00~ neutrophil~ 29/166    487/18~ 2.98e-16 5.11e-13 3.77e-13 CAMP/M~
      29
##  2 GO:00~ neutrophil~ 29/166    490/18~ 3.50e-16 5.11e-13 3.77e-13 CAMP/M~
      29
##  3 GO:00~ interferon~ 11/166    91/188~ 4.52e-10 4.40e- 7 3.24e- 7 PPARG/~
      11
##  4 GO:00~ lipid loca~ 20/166    440/18~ 2.03e- 9 1.48e- 6 1.09e- 6 PPARG/~
      20
##  5 GO:00~ reactive o~ 16/166    288/18~ 5.55e- 9 3.24e- 6 2.39e- 6 AIF1/C~
      16
```

```
##  6 GO:00~ response t~ 13/166    202/18~ 2.84e- 8 1.38e- 5 1.02e- 5 PPARG/~
      13
##  7 GO:19~ regulation~ 12/166    171/18~ 3.80e- 8 1.58e- 5 1.17e- 5 PPARG/~
      12
##  8 GO:00~ lipid tran~ 17/166    393/18~ 7.28e- 8 2.45e- 5 1.81e- 5 PPARG/~
      17
##  9 GO:00~ cellular r~ 12/166    182/18~ 7.56e- 8 2.45e- 5 1.81e- 5 PPARG/~
      12
## 10 GO:00~ regulation~ 13/166    240/18~ 2.13e- 7 6.18e- 5 4.55e- 5 IGFBP2~
      13
## # ... with 2,909 more rows
```

```
write.csv(result.enrichGO.de_ident0, file = "./Results_lists/enrichGOBP_DE_
    cluster1.csv")
```

3.2 Cluster 2 DE analysis and GO enrichment

The Cluster 2 mentioned in manuscript refers to cluster 1 in the codes following.

Calculate the DE genes.

```
de_ident1 <- FindMarkers(results, ident.1 = "1", only.pos = TRUE, logfc.
    threshold = 0.2)
head(de_ident1)
```

```
## # A tibble: 6 x 5
##   p_val avg_log2FC pct.1 pct.2 p_val_adj
##   <dbl>      <dbl> <dbl> <dbl>     <dbl>
## 1     0      0.421 0.848 0.69          0
## 2     0      0.400 0.776 0.555         0
## 3     0      0.387 0.274 0.097         0
## 4     0      0.345 0.976 0.977         0
## 5     0      0.369 0.272 0.094         0
## 6     0      0.556 0.416 0.133         0
```

Save data

```
write.csv(de_ident1, file = "./Results_lists/DE_cluster2.csv")
```

KEGG/GO enrichment:

```
de_entrez.ident1 <- bitr( geneID = rownames(de_ident1), fromType = "SYMBOL",
    toType = "ENTREZID", OrgDb = "org.Hs.eg.db", drop = TRUE ) $ ENTREZID
result.enrichKEGG.de_ident1 <- enrichKEGG(de_entrez.ident1, organism = "hsa",
    keyType = "ncbi-geneid")
result.enrichKEGG.de_ident1 <- replaceEntrezID(result.enrichKEGG.de_ident1)
```

```
result.enrichKEGG.de_ident1@result
```

```
## # A tibble: 200 x 9
##    ID       Description  GeneRatio BgRatio  pvalue p.adjust  qvalue geneID
    Count
```

```
##    <chr>    <chr>        <chr>     <chr>     <dbl>    <dbl>   <dbl> <chr>
   <int>
##  1 hsa04142 Lysosome     11/93     128/80~ 2.04e-7  4.08e-5 3.75e-5 DNASE~
      11
##  2 hsa04145 Phagosome    9/93      152/80~ 5.74e-5  5.74e-3 5.29e-3 NCF1/~
       9
##  3 hsa04966 Collecting ~ 4/93      27/8095 2.34e-4  1.19e-2 1.09e-2 ATP6V~
       4
##  4 hsa05110 Vibrio chol~ 5/93      50/8095 2.53e-4  1.19e-2 1.09e-2 ATP6A~
       5
##  5 hsa05171 Coronavirus~ 10/93     232/80~ 3.09e-4  1.19e-2 1.09e-2 CYBB/~
      10
##  6 hsa00030 Pentose pho~ 4/93      30/8095 3.56e-4  1.19e-2 1.09e-2 PGD/F~
       4
##  7 hsa03010 Ribosome     8/93      158/80~ 4.41e-4  1.26e-2 1.16e-2 RPL39~
       8
##  8 hsa05323 Rheumatoid ~ 6/93      93/8095 6.68e-4  1.67e-2 1.54e-2 CSF1/~
       6
##  9 hsa05120 Epithelial ~ 5/93      70/8095 1.21e-3  2.68e-2 2.47e-2 ATP6A~
       5
## 10 hsa04979 Cholesterol~ 4/93      50/8095 2.51e-3  5.02e-2 4.63e-2 LIPA/~
       4
## # ... with 190 more rows
```

```
write.csv(result.enrichKEGG.de_ident1, file = "./Results_lists/enrichKEGG_DE_
   cluster2.csv")
```

```
result.enrichGO.de_ident1 <- enrichGO(gene = de_entrez.ident1, OrgDb = "org.Hs
   .eg.db", ont = "BP")
result.enrichGO.de_ident1 <- replaceEntrezID(result.enrichGO.de_ident1)
result.enrichGO.de_ident1@result
```

```
## # A tibble: 2,543 x 9
##    ID     Description  GeneRatio BgRatio  pvalue p.adjust  qvalue geneID
   Count
##    <chr>  <chr>        <chr>     <chr>     <dbl>    <dbl>   <dbl> <chr>
   <int>
##  1 GO:00~ lipid catab~ 15/131    343/18~ 1.64e-8  4.17e-5 3.51e-5 FABP3/H~
      15
##  2 GO:00~ cell redox ~ 7/131     59/188~ 1.67e-7  1.52e-4 1.28e-4 PRDX1/N~
       7
##  3 GO:00~ neutrophil ~ 16/131    487/18~ 2.75e-7  1.52e-4 1.28e-4 FGR/HEX~
      16
##  4 GO:00~ bone resorp~ 7/131     64/188~ 2.96e-7  1.52e-4 1.28e-4 ATP6AP1~
       7
##  5 GO:00~ neutrophil ~ 16/131    490/18~ 2.99e-7  1.52e-4 1.28e-4 FGR/HEX~
      16
##  6 GO:00~ tissue remo~ 10/131    178/18~ 4.68e-7  1.98e-4 1.67e-4 ATP6AP1~
      10
##  7 GO:00~ translation~ 10/131    192/18~ 9.36e-7  3.40e-4 2.86e-4 RPL39/P~
      10
##  8 GO:00~ nuclear-tra~ 8/131     120/18~ 1.91e-6  5.97e-4 5.02e-4 RPL39/P~
       8
```

```
##  9 GO:00~ tissue home~ 11/131    261/18~ 2.11e-6  5.97e-4 5.02e-4 PRDX1/C~ 12
      11
## 10 GO:00~ positive re~ 6/131     58/188~ 3.01e-6  7.64e-4 6.42e-4 CSF1/TR~ 13
       6
## # ... with 2,533 more rows                                                 14
```

```
write.csv(result.enrichGO.de_ident1, file = "./Results_lists/enrichGOBP_DE_    1
   cluster2.csv")
```

3.3 Cluster 3 DE analysis

The Cluster 3 mentioned in manuscript refers to cluster 2 in the codes following.

Calculate the DE genes.

```
de_ident2 <- FindMarkers(results, ident.1 = "2", only.pos = TRUE, logfc.      1
   threshold = 0.2)
nrow(de_ident2)                                                               2
```

```
## [1] 660                                                                    1
```

Save data

```
write.csv(de_ident2, file = "./Results_lists/DE_cluster3.csv")                1
```

3.4 Cluster 4 DE analysis

The Cluster 4 mentioned in manuscript refers to cluster 3 in the codes following.

Calculate the DE genes.

```
de_ident3 <- FindMarkers(results, ident.1 = "3", only.pos = TRUE, logfc.      1
   threshold = 0.2)
head(de_ident3)                                                               2
```

```
## # A tibble: 6 x 5                                                         1
##   p_val avg_log2FC pct.1 pct.2 p_val_adj                                   2
##   <dbl>      <dbl> <dbl> <dbl>     <dbl>                                  3
## 1     0      0.244 0.274 0.028         0                                  4
## 2     0      0.208 0.175 0.002         0                                  5
## 3     0      2.34  0.983 0.684         0                                  6
## 4     0      0.770 0.473 0.014         0                                  7
## 5     0      0.319 0.225 0.009         0                                  8
## 6     0      0.885 0.352 0.003         0                                  9
```

Save data

```
write.csv(de_ident3, file = "./Results_lists/DE_cluster4.csv")                1
```

3.5 GO/KEGG enrichment and DE analysis in subpopulations of Cluster 3 after re-clustering

Load Cluster 3 object:

```
results.c2 <- readRDS("../4-Functional characterization of clustered
   populations/cluster2_clustered.seuratObject.Rds")
```

REMIND: the subpopulations cluster 1-4 (after reclustering the cluster 3) in the manuscript refer to the cluster 0-3 in the following codes.

Calculate the DE genes.

DE genes for each subpopulations:

```
de_subpop0.c2 <- FindMarkers(object = results.c2, ident.1 = "0", only.pos =
   TRUE, logfc.threshold = 0.2, verbose = FALSE)
de_subpop1.c2 <- FindMarkers(object = results.c2, ident.1 = "1", only.pos =
   TRUE, logfc.threshold = 0.2, verbose = FALSE)
de_subpop2.c2 <- FindMarkers(object = results.c2, ident.1 = "2", only.pos =
   TRUE, logfc.threshold = 0.2, verbose = FALSE)
de_subpop3.c2 <- FindMarkers(object = results.c2, ident.1 = "3", only.pos =
   TRUE, logfc.threshold = 0.2, verbose = FALSE)
```

Save data

```
write.csv(de_subpop0.c2, file = "./Results_lists/DE_subpop1_c3.csv")
write.csv(de_subpop1.c2, file = "./Results_lists/DE_subpop2_c3.csv")
write.csv(de_subpop2.c2, file = "./Results_lists/DE_subpop3_c3.csv")
write.csv(de_subpop3.c2, file = "./Results_lists/DE_subpop4_c3.csv")
```

Since the subpopulations 2 and 3 are very similar. We would like also to have the DE genes in both subpopulations 2 & 3 comparing to subpopulation 1 (we excluded cluster 4 because it's high similarity to DCs).

```
de_subpop1_2.c2 <- FindMarkers(object = results.c2, ident.1 = c("1", "2"),
   ident.2 = ("0"), only.pos = TRUE, logfc.threshold = 0.2, verbose = FALSE)
write.csv(de_subpop1_2.c2, file = "./Results_lists/DE_subpop2_and_3_of_c3.csv"
   )
```

KEGG enrichment:

```
de_entrez.subpop0.c2 <- bitr( geneID = rownames(de_subpop0.c2), fromType = "
   SYMBOL", toType = "ENTREZID", OrgDb = "org.Hs.eg.db", drop = TRUE ) $
   ENTREZID
de_entrez.subpop1.c2 <- bitr( geneID = rownames(de_subpop1.c2), fromType = "
   SYMBOL", toType = "ENTREZID", OrgDb = "org.Hs.eg.db", drop = TRUE ) $
   ENTREZID
de_entrez.subpop2.c2 <- bitr( geneID = rownames(de_subpop2.c2), fromType = "
   SYMBOL", toType = "ENTREZID", OrgDb = "org.Hs.eg.db", drop = TRUE ) $
   ENTREZID
de_entrez.subpop3.c2 <- bitr( geneID = rownames(de_subpop3.c2), fromType = "
   SYMBOL", toType = "ENTREZID", OrgDb = "org.Hs.eg.db", drop = TRUE ) $
   ENTREZID
de_entrez.subpop1_2.c2 <- bitr( geneID = rownames(de_subpop1_2.c2), fromType =
    "SYMBOL", toType = "ENTREZID", OrgDb = "org.Hs.eg.db", drop = TRUE ) $
   ENTREZID
```

```
                                                                                                    6
result.enrichKEGG.de_entrez.subpop0.c2 <- enrichKEGG(de_entrez.subpop0.c2,                          7
    organism = "hsa", keyType = "ncbi-geneid")
result.enrichKEGG.de_entrez.subpop1.c2 <- enrichKEGG(de_entrez.subpop1.c2,                          8
    organism = "hsa", keyType = "ncbi-geneid")
result.enrichKEGG.de_entrez.subpop2.c2 <- enrichKEGG(de_entrez.subpop2.c2,                          9
    organism = "hsa", keyType = "ncbi-geneid")
result.enrichKEGG.de_entrez.subpop3.c2 <- enrichKEGG(de_entrez.subpop3.c2,                          10
    organism = "hsa", keyType = "ncbi-geneid")
result.enrichKEGG.de_entrez.subpop1_2.c2 <- enrichKEGG(de_entrez.subpop1_2.c2,                      11
     organism = "hsa", keyType = "ncbi-geneid")
                                                                                                    12
result.enrichGO.de_entrez.subpop0.c2 <- enrichGO(gene = de_entrez.subpop0.c2,                       13
    OrgDb = "org.Hs.eg.db", ont = "BP")
result.enrichGO.de_entrez.subpop1.c2 <- enrichGO(gene = de_entrez.subpop1.c2,                       14
    OrgDb = "org.Hs.eg.db", ont = "BP")
result.enrichGO.de_entrez.subpop2.c2 <- enrichGO(gene = de_entrez.subpop2.c2,                       15
    OrgDb = "org.Hs.eg.db", ont = "BP")
result.enrichGO.de_entrez.subpop3.c2 <- enrichGO(gene = de_entrez.subpop3.c2,                       16
    OrgDb = "org.Hs.eg.db", ont = "BP")
result.enrichGO.de_entrez.subpop1_2.c2 <- enrichGO(gene = de_entrez.subpop1_2.                      17
    c2, OrgDb = "org.Hs.eg.db", ont = "BP")
                                                                                                    18
result.enrichKEGG.de_entrez.subpop0.c2 <- replaceEntrezID(result.enrichKEGG.de                      19
    _entrez.subpop0.c2)
result.enrichKEGG.de_entrez.subpop1.c2 <- replaceEntrezID(result.enrichKEGG.de                      20
    _entrez.subpop1.c2)
result.enrichKEGG.de_entrez.subpop2.c2 <- replaceEntrezID(result.enrichKEGG.de                      21
    _entrez.subpop2.c2)
result.enrichKEGG.de_entrez.subpop3.c2 <- replaceEntrezID(result.enrichKEGG.de                      22
    _entrez.subpop3.c2)
result.enrichKEGG.de_entrez.subpop1_2.c2 <- replaceEntrezID(result.enrichKEGG.                      23
    de_entrez.subpop1_2.c2)
                                                                                                    24
result.enrichGO.de_entrez.subpop0.c2 <- replaceEntrezID(result.enrichGO.de_                         25
    entrez.subpop0.c2)
result.enrichGO.de_entrez.subpop1.c2 <- replaceEntrezID(result.enrichGO.de_                         26
    entrez.subpop1.c2)
result.enrichGO.de_entrez.subpop2.c2 <- replaceEntrezID(result.enrichGO.de_                         27
    entrez.subpop2.c2)
result.enrichGO.de_entrez.subpop3.c2 <- replaceEntrezID(result.enrichGO.de_                         28
    entrez.subpop3.c2)
result.enrichGO.de_entrez.subpop1_2.c2 <- replaceEntrezID(result.enrichGO.de_                       29
    entrez.subpop1_2.c2)
```

```
write.csv(result.enrichKEGG.de_entrez.subpop0.c2, file = "./Results_lists/                          1
    enrichKEGG_DE_subpop1_c3.csv")
write.csv(result.enrichKEGG.de_entrez.subpop1.c2, file = "./Results_lists/                          2
    enrichKEGG_DE_subpop2_c3.csv")
write.csv(result.enrichKEGG.de_entrez.subpop2.c2, file = "./Results_lists/                          3
    enrichKEGG_DE_subpop3_c3.csv")
write.csv(result.enrichKEGG.de_entrez.subpop3.c2, file = "./Results_lists/                          4
    enrichKEGG_DE_subpop4_c3.csv")
write.csv(result.enrichKEGG.de_entrez.subpop1_2.c2, file = "./Results_lists/                        5
```

```
    enrichKEGG_DE_subpop2_3.c3.csv")
                                                                                6
write.csv(result.enrichGO.de_entrez.subpop0.c2, file = "./Results_lists/        7
    enrichGOBP_DE_subpop1_c3.csv")
write.csv(result.enrichGO.de_entrez.subpop1.c2, file = "./Results_lists/        8
    enrichGOBP_DE_subpop2_c3.csv")
write.csv(result.enrichGO.de_entrez.subpop2.c2, file = "./Results_lists/        9
    enrichGOBP_DE_subpop3_c3.csv")
write.csv(result.enrichGO.de_entrez.subpop3.c2, file = "./Results_lists/        10
    enrichGOBP_DE_subpop4_c3.csv")
write.csv(result.enrichGO.de_entrez.subpop1_2.c2, file = "./Results_lists/      11
    enrichGOBP_DE_subpop2_3_C3.csv")
```

Head 10 lines in GO:

For subpop 1:

```
head(result.enrichGO.de_entrez.subpop0.c2@result, 10)                           1
```

```
## # A tibble: 10 x 9                                                           1
##    ID     Description GeneRatio BgRatio   pvalue p.adjust   qvalue geneID   2
    Count
##    <chr>  <chr>       <chr>     <chr>      <dbl>    <dbl>    <dbl> <chr>    3
    <int>
##  1 GO:00~ neutrophil~ 91/749    487/18~ 3.27e-36 1.38e-32 1.07e-32 CYBB/M~   4
      91
##  2 GO:00~ neutrophil~ 91/749    490/18~ 5.45e-36 1.38e-32 1.07e-32 CYBB/M~   5
      91
##  3 GO:00~ oxidative ~ 37/749    149/18~ 1.25e-19 2.12e-16 1.63e-16 RHOA/U~   6
      37
##  4 GO:00~ ATP metabo~ 47/749    311/18~ 2.51e-15 3.19e-12 2.45e-12 FBP1/T~   7
      47
##  5 GO:00~ phagocytos~ 52/749    382/18~ 6.65e-15 6.76e-12 5.21e-12 ITGAM/~   8
      52
##  6 GO:00~ electron t~ 33/749    184/18~ 2.93e-13 2.48e-10 1.91e-10 CYBB/S~   9
      33
##  7 GO:00~ mitochondr~ 23/749    97/188~ 2.97e-12 2.16e- 9 1.66e- 9 UQCRH/~   10
      23
##  8 GO:00~ ATP synthe~ 23/749    98/188~ 3.74e-12 2.38e- 9 1.83e- 9 UQCRH/~   11
      23
##  9 GO:00~ regulation~ 45/749    360/18~ 9.28e-12 5.24e- 9 4.04e- 9 RGCC/S~   12
      45
## 10 GO:00~ cellular r~ 31/749    187/18~ 1.31e-11 6.66e- 9 5.13e- 9 UQCRH/~   13
      31
```

For subpop 2:

```
head(result.enrichGO.de_entrez.subpop1.c2@result, 10)                           1
```

```
## # A tibble: 10 x 9                                                           1
##    ID     Description GeneRatio BgRatio   pvalue p.adjust   qvalue geneID   2
    Count
##    <chr>  <chr>       <chr>     <chr>      <dbl>    <dbl>    <dbl> <chr>    3
    <int>
```

```
##  1 GO:00~ neutrophil~ 39/299    487/18~ 7.13e-17 1.58e-13 1.21e-13 MME/FT~  4
      39
##  2 GO:00~ neutrophil~ 39/299    490/18~ 8.78e-17 1.58e-13 1.21e-13 MME/FT~  5
      39
##  3 GO:00~ cellular r~ 21/299    182/18~ 1.35e-12 1.63e- 9 1.25e- 9 CCL18/~  6
      21
##  4 GO:00~ response t~ 21/299    202/18~ 1.03e-11 9.28e- 9 7.11e- 9 CCL18/~  7
      21
##  5 GO:00~ neutrophil~ 14/299    103/18~ 9.05e-10 6.53e- 7 5.01e- 7 CCL18/~  8
      14
##  6 GO:00~ response t~ 20/299    250/18~ 3.43e- 9 1.81e- 6 1.39e- 6 NUPR1/~  9
      20
##  7 GO:19~ cellular d~ 14/299    114/18~ 3.52e- 9 1.81e- 6 1.39e- 6 SOD2/A~  10
      14
##  8 GO:00~ detoxifica~ 15/299    138/18~ 5.33e- 9 2.40e- 6 1.84e- 6 SOD2/M~  11
      15
##  9 GO:00~ cellular r~ 14/299    122/18~ 8.59e- 9 3.01e- 6 2.31e- 6 SOD2/A~  12
      14
## 10 GO:19~ neutrophil~ 14/299    122/18~ 8.59e- 9 3.01e- 6 2.31e- 6 CCL18/~  13
      14
```

For subpop 3:

```
head(result.enrichGO.de_entrez.subpop2.c2@result, 10)                          1
```

```
## # A tibble: 10 x 9                                                          1
##    ID     Description GeneRatio BgRatio   pvalue p.adjust   qvalue geneID   2
   Count
##    <chr>  <chr>       <chr>     <chr>      <dbl>    <dbl>    <dbl> <chr>    3
   <int>
##  1 GO:00~ neutrophil~ 70/526    487/18~ 3.13e-30 1.07e-26 7.63e-27 CTSB/P~  4
      70
##  2 GO:00~ neutrophil~ 70/526    490/18~ 4.62e-30 1.07e-26 7.63e-27 CTSB/P~  5
      70
##  3 GO:00~ positive r~ 41/526    447/18~ 2.37e-11 3.67e- 8 2.61e- 8 HIF1A/~  6
      41
##  4 GO:00~ regulation~ 27/526    212/18~ 4.77e-11 5.54e- 8 3.94e- 8 LGMN/A~  7
      27
##  5 GO:19~ positive r~ 18/526    94/188~ 9.84e-11 9.14e- 8 6.51e- 8 SPP1/A~  8
      18
##  6 GO:00~ maintenanc~ 33/526    324/18~ 1.49e-10 1.10e- 7 7.85e- 8 ABCA1/~  9
      33
##  7 GO:00~ cellular r~ 26/526    208/18~ 1.66e-10 1.10e- 7 7.85e- 8 ABCA1/~  10
      26
##  8 GO:00~ response t~ 33/526    334/18~ 3.31e-10 1.92e- 7 1.37e- 7 ABCA1/~  11
      33
##  9 GO:00~ cellular r~ 26/526    222/18~ 6.99e-10 3.25e- 7 2.31e- 7 ABCA1/~  12
      26
## 10 GO:00~ myeloid le~ 26/526    222/18~ 6.99e-10 3.25e- 7 2.31e- 7 LGMN/A~  13
      26
```

For subpop 4:

```
head(result.enrichGO.de_entrez.subpop3.c2@result, 10)                          1
```

```
## # A tibble: 10 x 9                                                                                    1
##    ID     Description GeneRatio BgRatio   pvalue p.adjust   qvalue geneID   2
   Count
##    <chr>  <chr>       <chr>     <chr>      <dbl>    <dbl>    <dbl> <chr>    3
   <int>
##  1 GO:00~ SRP-depend~ 79/926    105/18~ 1.35e-81 6.99e-78 5.90e-78 RPL18/~   4
      79
##  2 GO:00~ cotranslat~ 80/926    109/18~ 4.03e-81 1.05e-77 8.85e-78 RPL18/~   5
      80
##  3 GO:00~ protein ta~ 81/926    120/18~ 2.86e-77 4.95e-74 4.18e-74 RPL18/~   6
      81
##  4 GO:00~ establishm~ 81/926    124/18~ 1.81e-75 2.35e-72 1.99e-72 RPL18/~   7
      81
##  5 GO:00~ translatio~ 97/926    192/18~ 3.70e-75 3.84e-72 3.25e-72 RPL18/~   8
      97
##  6 GO:00~ nuclear-tr~ 79/926    120/18~ 5.07e-74 4.38e-71 3.70e-71 RPL18/~   9
      79
##  7 GO:00~ protein lo~ 84/926    152/18~ 1.97e-69 1.46e-66 1.23e-66 RPL18/~   10
      84
##  8 GO:00~ nuclear-tr~ 90/926    210/18~ 1.02e-61 6.60e-59 5.58e-59 RPL18/~   11
      90
##  9 GO:00~ viral gene~ 87/926    195/18~ 1.69e-61 9.77e-59 8.25e-59 RPL18/~   12
      87
## 10 GO:00~ viral tran~ 81/926    178/18~ 4.02e-58 2.09e-55 1.76e-55 RPL18/~   13
      81
```

For subpop 2,3:

```
head(result.enrichGO.de_entrez.subpop1_2.c2@result, 10)                                                1
```

```
## # A tibble: 10 x 9                                                                                    1
##    ID     Description GeneRatio BgRatio   pvalue p.adjust   qvalue geneID   2
   Count
##    <chr>  <chr>       <chr>     <chr>      <dbl>    <dbl>    <dbl> <chr>    3
   <int>
##  1 GO:00~ neutrophil~ 51/353    487/18~ 8.26e-24 2.24e-20 1.57e-20 PSAP/C~   4
      51
##  2 GO:00~ neutrophil~ 51/353    490/18~ 1.09e-23 2.24e-20 1.57e-20 PSAP/C~   5
      51
##  3 GO:19~ positive r~ 18/353    94/188~ 1.27e-13 1.73e-10 1.22e-10 PLTP/A~   6
      18
##  4 GO:00~ regulation~ 32/353    425/18~ 2.48e-11 2.54e- 8 1.78e- 8 MGLL/N~   7
      32
##  5 GO:00~ response t~ 22/353    202/18~ 3.36e-11 2.75e- 8 1.93e- 8 NR1H3/~   8
      22
##  6 GO:00~ regulation~ 33/353    470/18~ 7.61e-11 4.47e- 8 3.14e- 8 SERPIN~   9
      33
##  7 GO:00~ lipid tran~ 30/353    393/18~ 7.65e-11 4.47e- 8 3.14e- 8 PSAP/P~   10
      30
##  8 GO:00~ cholestero~ 13/353    64/188~ 1.54e-10 7.86e- 8 5.52e- 8 PLTP/A~   11
      13
##  9 GO:00~ cellular r~ 20/353    182/18~ 2.28e-10 9.99e- 8 7.02e- 8 NR1H3/~   12
      20
```

```
## 10 GO:00~ cellular r~ 23/353    246/18~ 2.64e-10 9.99e- 8 7.02e- 8 ABCA1/~
      23
```

4 Session information

R sesssion:

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.3 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/liblapack.so.3
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_GB.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_GB.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_GB.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] parallel  stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] org.Hs.eg.db_3.12.0    AnnotationDbi_1.52.0   IRanges_2.24.1
##  [4] S4Vectors_0.28.1       Biobase_2.50.0         BiocGenerics_0.36.1
##  [7] clusterProfiler_3.18.1 dplyr_1.0.7            ggplot2_3.3.5
## [10] SeuratObject_4.0.2     Seurat_4.0.3
##
## loaded via a namespace (and not attached):
##   [1] shadowtext_0.0.8      fastmatch_1.1-3       plyr_1.8.6
##   [4] igraph_1.2.6          lazyeval_0.2.2        splines_4.0.3
##   [7] BiocParallel_1.24.1   listenv_0.8.0         scattermore_0.7
##  [10] digest_0.6.27         htmltools_0.5.1.1     GOSemSim_2.16.1
##  [13] viridis_0.6.1         GO.db_3.12.1          fansi_0.5.0
##  [16] magrittr_2.0.1        memoise_2.0.0         tensor_1.5
##  [19] cluster_2.1.0         ROCR_1.0-11           limma_3.46.0
##  [22] graphlayouts_0.7.1    globals_0.14.0        matrixStats_0.60.0
##  [25] spatstat.sparse_2.0-0 enrichplot_1.10.2     colorspace_2.0-2
##  [28] blob_1.2.2            ggrepel_0.9.1         xfun_0.24
##  [31] crayon_1.4.1          jsonlite_1.7.2        scatterpie_0.1.6
##  [34] spatstat.data_2.1-0   survival_3.2-7        zoo_1.8-9
##  [37] glue_1.4.2            polyclip_1.10-0       gtable_0.3.0
##  [40] leiden_0.3.9          future.apply_1.7.0    abind_1.4-5
##  [43] scales_1.1.1          DOSE_3.16.0           DBI_1.1.1
##  [46] miniUI_0.1.1.1        Rcpp_1.0.7            viridisLite_0.4.0
##  [49] xtable_1.8-4          reticulate_1.20       spatstat.core_2.3-0
```

```
##  [52] bit_4.0.4              htmlwidgets_1.5.3      httr_1.4.2
##  [55] fgsea_1.16.0           RColorBrewer_1.1-2     ellipsis_0.3.2
##  [58] ica_1.0-2              farver_2.1.0           pkgconfig_2.0.3
##  [61] uwot_0.1.10.9000       deldir_0.2-10          utf8_1.2.2
##  [64] tidyselect_1.1.1       rlang_0.4.11           reshape2_1.4.4
##  [67] later_1.2.0            munsell_0.5.0          tools_4.0.3
##  [70] cachem_1.0.5           downloader_0.4         cli_3.0.1
##  [73] generics_0.1.0         RSQLite_2.2.7          ggridges_0.5.3
##  [76] evaluate_0.14          stringr_1.4.0          fastmap_1.1.0
##  [79] yaml_2.2.1             goftest_1.2-2          knitr_1.33
##  [82] bit64_4.0.5            fitdistrplus_1.1-5     tidygraph_1.2.0
##  [85] purrr_0.3.4            RANN_2.6.1             ggraph_2.0.5
##  [88] pbapply_1.4-3          future_1.21.0          nlme_3.1-152
##  [91] mime_0.11              DO.db_2.9              compiler_4.0.3
##  [94] rstudioapi_0.13        plotly_4.9.4.1         png_0.1-7
##  [97] spatstat.utils_2.2-0   tibble_3.1.3           tweenr_1.0.2
## [100] stringi_1.7.3          lattice_0.20-41        Matrix_1.3-4
## [103] vctrs_0.3.8            pillar_1.6.2           lifecycle_1.0.0
## [106] BiocManager_1.30.16    spatstat.geom_2.2-2    lmtest_0.9-38
## [109] RcppAnnoy_0.0.19       data.table_1.14.0      cowplot_1.1.1
## [112] irlba_2.3.3            httpuv_1.6.1           patchwork_1.1.1
## [115] qvalue_2.22.0          R6_2.5.0               promises_1.2.0.1
## [118] KernSmooth_2.23-20     gridExtra_2.3          parallelly_1.27.0
## [121] codetools_0.2-18       MASS_7.3-53            assertthat_0.2.1
## [124] withr_2.4.2            sctransform_0.3.2      mgcv_1.8-33
## [127] grid_4.0.3             rpart_4.1-15           tidyr_1.1.3
## [130] rvcheck_0.1.8          rmarkdown_2.9          Rtsne_0.15
## [133] ggforce_0.3.3          shiny_1.6.0
```

References

1. Hao Y, Hao S, Andersen-Nissen E, Mauck III WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zagar M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LB, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. *Cell* [Internet] 2021; Available from: https://doi.org/10.1016/j.cell.2021.04.048.

2. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: An r package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* 2012; 16: 284–287.