DE analysis

Shuai Guo, Yunshan Duan

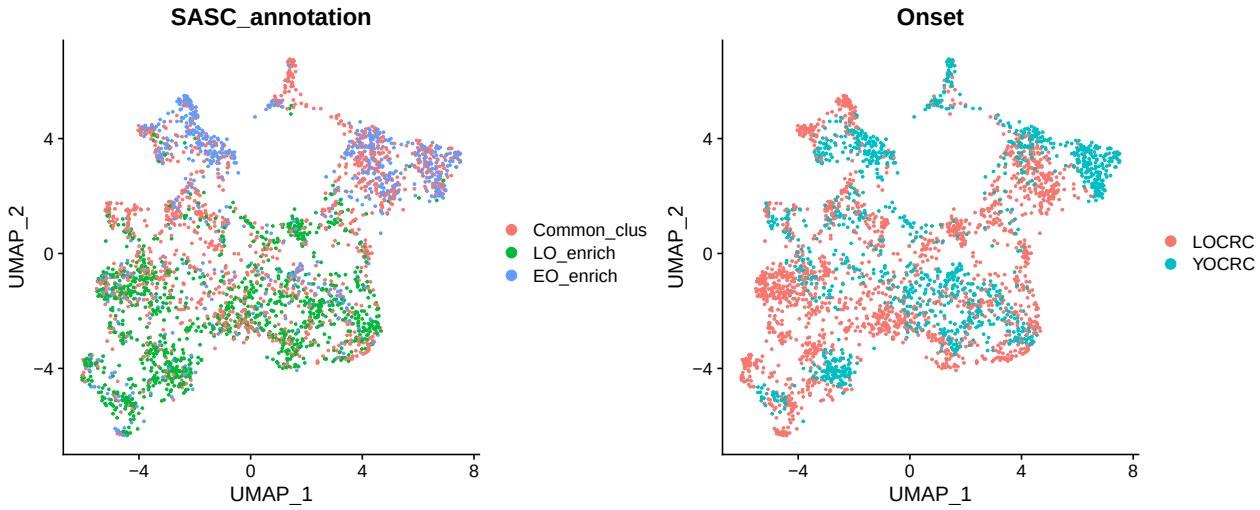
2024-03-25

```
# load the pacakges
  library(glmpca)
  library(ggplot2); theme_set(theme_bw())
  library(cluster)
  library(salso)
  library(ggpubr)
  library(dplyr)
  library(fdrtool)
  library(Seurat)
# read the model output
  output = readRDS(file = "../output/realdata/data/SASC_output.rds")
  clusterinfo = readRDS(file = "../output/realdata/data/clusterinfo.rds")
```

```
# print the clustering results
  output$SASC_cluster = factor(output$SASC_cluster, c(1,6,2:5,7,8,10,9,11))
  table(output$SASC_cluster, output$annotation_final)
```

```
##
##      cd4T_hp cd4T_other cd4T_rg CD8T_em CD8T_ex CD8T_other
##   1        8         12     160       0       0          0
##   6      122         22     279       0       0          0
##   2      208         48      30       0       0          0
##   3       95         13      91       0       0          0
##   4      405         13       6       0       0          0
##   5       20         90       4       0       0          0
##   7        0          0       0     273      72         28
##   8        0          0       0      67      56         54
##   10       0          0       0      26     189          3
##   9        0          0       0     271      12        339
##   11       0          0       0      20      33         70
```

```
# Assign annotation results
  SASC_annotation = rep("Common_clus", dim(output)[2])
  SASC_annotation[which(output$SASC_cluster == 4 | output$SASC_cluster == 5 | output$SASC_cluster == 9)]
  SASC_annotation[which(output$SASC_cluster == 6 | output$SASC_cluster == 10| output$SASC_cluster ==11)]
  output$SASC_annotation = factor(SASC_annotation, levels = c("Common_clus", "LO_enrich", "EO_enrich"))
# Visualize the results
  DimPlot(output, group.by = "SASC_annotation")+
    DimPlot(output, group.by = "Onset")
```




```
  Idents(output) = output$T_cell_type
  CD4 = subset(output, idents = "CD4")
  CD8 = subset(output, idents = "CD8")
```

The CD4 T helper cells

```
# Identify conditional specific genes for CD4 subtypes
  Idents(CD4) = CD4$annotation_final
  CD4_help = subset(CD4, idents = "cd4T_hp")
  Idents(CD4_help) = CD4_help$SASC_annotation
  output.markers <- FindAllMarkers(CD4_help, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25)
```

```
## Calculating cluster Common_clus
```

```
## Calculating cluster LO_enrich
```

```
## Calculating cluster EO_enrich
```

```
  top2 = output.markers %>%
      group_by(cluster) %>%
      slice_max(n = 5, order_by = avg_log2FC)
  top2
```

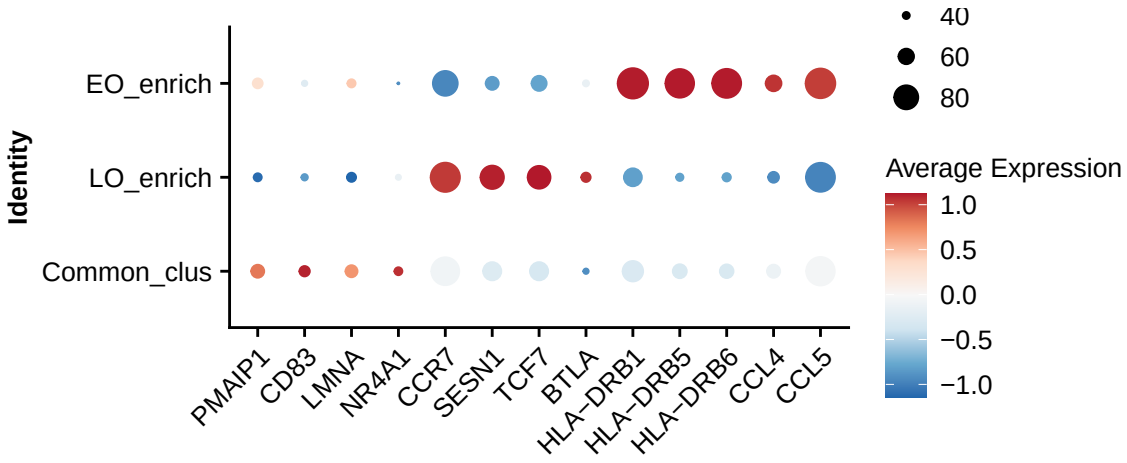
```
## # A tibble: 15 x 7
## # Groups:   cluster [3]
##       p_val avg_log2FC pct.1 pct.2 p_val_adj cluster     gene
##       <dbl>      <dbl> <dbl> <dbl>     <dbl> <fct>       <chr>
##  1 4.59e- 7      0.882 0.54  0.433  5.75e- 3 Common_clus PMAIP1
##  2 3.00e- 4      0.813 0.479 0.38   1    e+ 0 Common_clus CD83
##  3 1.19e- 3      0.749 0.518 0.444  1    e+ 0 Common_clus LMNA
##  4 1.40e- 3      0.647 0.421 0.338  1    e+ 0 Common_clus NR4A1
##  5 8.23e- 3      0.542 0.907 0.921  1    e+ 0 Common_clus CCL5
##  6 9.74e- 8      0.662 0.925 0.871  1.22e- 3 LO_enrich   CCR7
##  7 7.12e- 3      0.559 0.4   0.335  1    e+ 0 LO_enrich   ARHGAP5
##  8 4.06e- 6      0.526 0.786 0.628  5.08e- 2 LO_enrich   SESN1
##  9 1.49e- 8      0.510 0.772 0.642  1.87e- 4 LO_enrich   TCF7
## 10 1.11e- 3      0.506 0.454 0.365  1    e+ 0 LO_enrich   BTLA
## 11 6.65e-39      2.04  0.951 0.682  8.33e-35 EO_enrich   HLA-DRB1
```

```
## 12 2.04e-14      1.78  0.738 0.51   2.55e-10 EO_enrich   HLA-DRA
## 13 4.90e-42      1.61  0.91  0.473  6.14e-38 EO_enrich   HLA-DRB5
## 14 3.07e-42      1.55  0.918 0.484  3.85e-38 EO_enrich   HLA-DRB6
## 15 6.34e- 4      1.52  0.607 0.512  1    e+ 0 EO_enrich   CCL4
```

```
# Print our the final DE genes that we are interested in.
  final_list=c("PMAIP1","CD83","LMNA","NR4A1","CCR7","SESN1","TCF7","BTLA","HLA-DRB1","HLA-DRB5","HLA-DI
  p =
  DotPlot(CD4_help, features = unique(final_list), dot.scale=5, cols="RdBu") +
        theme(title=element_text(size=10), axis.text.x=element_text(size=10, angle=45, hjust=1),
              axis.title.x=element_text(size=0), axis.text.y=element_text(size=10),
              axis.title.y=element_text(size=10, face="bold"), legend.position = "right",
              legend.text=element_text(size=10), legend.title=element_text(size=10))
```

```
## Warning: Scaling data with a low number of groups may produce misleading
## results
```

```
  p
```



The CD4 Treg cells

```
# Subset the Tregs
  CD4_reg = subset(CD4, idents = "cd4T_rg")
  Idents(CD4_reg) = CD4_reg$SASC_annotation
  output.markers <- FindAllMarkers(CD4_reg, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25)
```

```
## Calculating cluster Common_clus
```

```
## Calculating cluster LO_enrich
```

```
## Calculating cluster EO_enrich
```

```
  top2 = output.markers %>%
      group_by(cluster) %>%
      slice_max(n = 4, order_by = avg_log2FC)
  top2
```

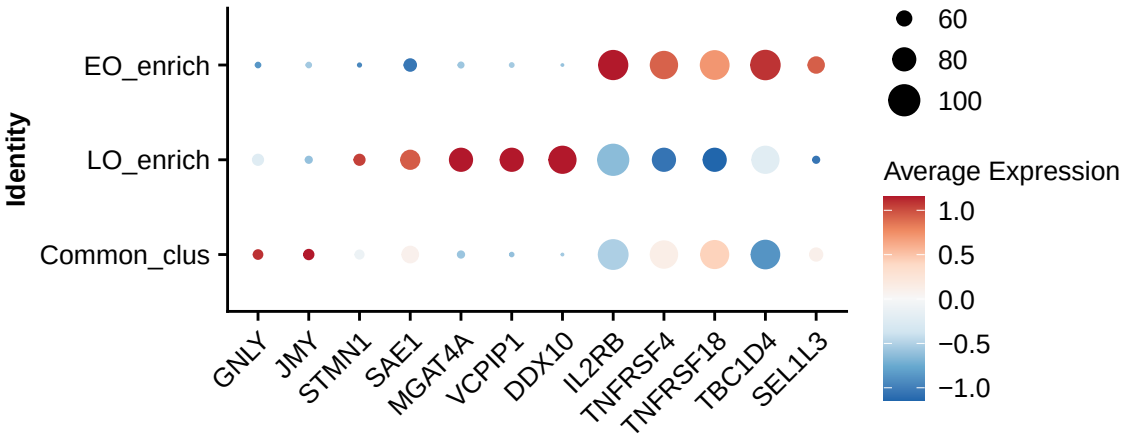
```
## # A tibble: 12 x 7
## # Groups:   cluster [3]
##         p_val avg_log2FC pct.1 pct.2 p_val_adj cluster     gene
##         <dbl>      <dbl> <dbl> <dbl>     <dbl> <fct>       <chr>
##  1 0.000100        1.27  0.448 0.332     1     Common_clus STMN1
```

```
##  2 0.00854         1.13  0.463 0.367    1      Common_clus GNLY
##  3 0.000614        0.534 0.48  0.363    1      Common_clus JMY
##  4 0.00416         0.521 0.377 0.284    1      Common_clus NUSAP1
##  5 0.00197         4.26  0.4   0.107    1      LO_enrich   IL17A
##  6 0.000211        1.65  0.8   0.388    1      LO_enrich   MGAT4A
##  7 0.000231        1.58  0.8   0.338    1      LO_enrich   VCPIP1
##  8 0.00000465      1.43  0.9   0.291    0.0583 LO_enrich   DDX10
##  9 0.00282         0.543 0.961 0.942    1      EO_enrich   LTB
## 10 0.0000482       0.444 0.957 0.935    0.603  EO_enrich   TBC1D4
## 11 0.00501         0.359 0.631 0.546    1      EO_enrich   SEL1L3
## 12 0.000182        0.347 0.95  0.966    1      EO_enrich   IL2RB
```

```
# Print our the final DE genes that we are interested in.
  final_list=c("GNLY","JMY","STMN1","SAE1","MGAT4A","VCPIP1","DDX10","IL2RB","TNFRSF4","TNFRSF18","TBC1
  p =
  DotPlot(CD4_reg, features = final_list, dot.scale=5, cols="RdBu") +
        theme(title=element_text(size=10), axis.text.x=element_text(size=10, angle=45, hjust=1),
              axis.title.x=element_text(size=0), axis.text.y=element_text(size=10),
              axis.title.y=element_text(size=10, face="bold"), legend.position = "right",
              legend.text=element_text(size=10), legend.title=element_text(size=10))
```

```
## Warning: Scaling data with a low number of groups may produce misleading
## results
```

```
  p
```



The CD8 T effector memory cells

```
# subset cell type
  Idents(CD8) = CD8$annotation_final
  CD8_em = subset(CD8, idents = "CD8T_em")
  Idents(CD8_em) = CD8_em$SASC_annotation
  output.markers <- FindAllMarkers(CD8_em, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25)
```

```
## Calculating cluster Common_clus
```

```
## Calculating cluster LO_enrich
```

```
## Calculating cluster EO_enrich
```

```
  top2 = output.markers %>%
      group_by(cluster) %>%
```

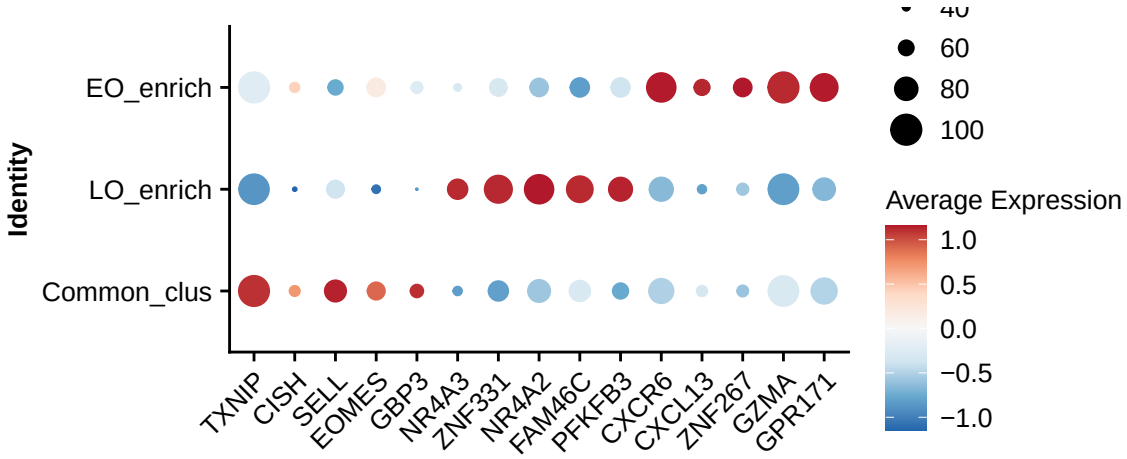
```
    slice_max(n = 5, order_by = avg_log2FC)
top2
```

```
## # A tibble: 15 x 7
## # Groups:   cluster [3]
##       p_val avg_log2FC pct.1 pct.2 p_val_adj cluster     gene
##       <dbl>      <dbl> <dbl> <dbl>     <dbl> <fct>       <chr>
##  1 2.21e- 7      0.923 0.997 0.987  2.76e- 3 Common_clus TXNIP
##  2 1.08e- 4      0.856 0.474 0.328  1    e+ 0 Common_clus CISH
##  3 9.83e- 5      0.717 0.762 0.647  1    e+ 0 Common_clus SELL
##  4 2.46e- 9      0.704 0.659 0.451  3.08e- 5 Common_clus EOMES
##  5 2.18e-11      0.641 0.538 0.29   2.72e- 7 Common_clus GBP3
##  6 7.99e-17      1.01  0.72  0.427  1.00e-12 LO_enrich   NR4A3
##  7 4.66e-18      0.965 0.919 0.71   5.84e-14 LO_enrich   ZNF331
##  8 2.25e-18      0.932 0.956 0.772  2.82e-14 LO_enrich   NR4A2
##  9 6.76e-12      0.900 0.886 0.738  8.47e- 8 LO_enrich   FAM46C
## 10 1.06e-11      0.872 0.815 0.622  1.33e- 7 LO_enrich   PFKFB3
## 11 1.54e- 5      1.05  0.957 0.833  1.92e- 1 EO_enrich   CXCR6
## 12 5.17e- 3      0.994 0.609 0.458  1    e+ 0 EO_enrich   CXCL13
## 13 8.79e- 4      0.982 0.674 0.501  1    e+ 0 EO_enrich   ZNF267
## 14 1.26e- 5      0.977 1     0.989  1.58e- 1 EO_enrich   GZMA
## 15 4.09e- 4      0.891 0.913 0.83   1    e+ 0 EO_enrich   GPR171
```

```
# Print our the final DE genes that we are interested in.
  p =
  DotPlot(CD8_em, features = unique(top2$gene), dot.scale=5, cols="RdBu") +
        theme(title=element_text(size=10), axis.text.x=element_text(size=10, angle=45, hjust=1),
              axis.title.x=element_text(size=0), axis.text.y=element_text(size=10),
              axis.title.y=element_text(size=10, face="bold"), legend.position = "right",
              legend.text=element_text(size=10), legend.title=element_text(size=10))
```

```
## Warning: Scaling data with a low number of groups may produce misleading
## results
```

```
  p
```



The CD8 T exhausted cells

```r
# subset cell type
  CD8_ex = subset(CD8, idents = "CD8T_ex")
  Idents(CD8_ex) = CD8_ex$SASC_annotation
  output.markers <- FindAllMarkers(CD8_ex, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25)
```

```
## Calculating cluster Common_clus
```

```
## Calculating cluster LO_enrich
```

```
## Calculating cluster EO_enrich
```

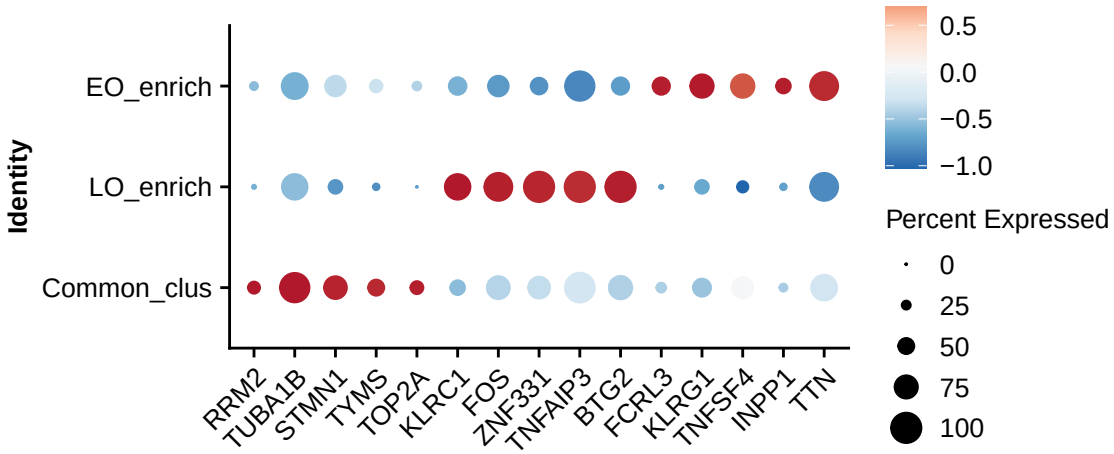
```r
  top2 = output.markers %>%
      group_by(cluster) %>%
      slice_max(n = 5, order_by = avg_log2FC)
  top2
```

```
## # A tibble: 15 x 7
## # Groups:   cluster [3]
##             p_val avg_log2FC pct.1 pct.2 p_val_adj cluster     gene
##             <dbl>      <dbl> <dbl> <dbl>     <dbl> <fct>       <chr>
##  1 0.0000794          2.24  0.359 0.209 0.995     Common_clus RRM2
##  2 0.000000129        1.91  0.969 0.846 0.00162   Common_clus TUBA1B
##  3 0.000249           1.69  0.734 0.645 1         Common_clus STMN1
##  4 0.000758           1.20  0.5   0.368 1         Common_clus TYMS
##  5 0.0000511          1.17  0.391 0.231 0.640     Common_clus TOP2A
##  6 0.000353           2.18  0.833 0.517 1         LO_enrich   KLRC1
##  7 0.000298           2.05  0.917 0.689 1         LO_enrich   FOS
##  8 0.0000430          1.62  1     0.589 0.538     LO_enrich   ZNF331
##  9 0.00000721         1.55  1     0.974 0.0903    LO_enrich   TNFAIP3
## 10 0.0000140          1.53  1     0.623 0.175     LO_enrich   BTG2
## 11 0.00000000743      1.14  0.545 0.264 0.0000931 EO_enrich   FCRL3
## 12 0.0000299          0.979 0.757 0.557 0.374     EO_enrich   KLRG1
## 13 0.000164           0.817 0.761 0.65  1         EO_enrich   TNFSF4
## 14 0.000000669        0.784 0.45  0.214 0.00838   EO_enrich   INPP1
## 15 0.0000218          0.718 0.919 0.85  0.273     EO_enrich   TTN
```

```r
  p =
  DotPlot(CD8_ex, features = unique(top2$gene), dot.scale=5, cols="RdBu") +
        theme(title=element_text(size=10), axis.text.x=element_text(size=10, angle=45, hjust=1),
              axis.title.x=element_text(size=0), axis.text.y=element_text(size=10),
              axis.title.y=element_text(size=10, face="bold"), legend.position = "right",
              legend.text=element_text(size=10), legend.title=element_text(size=10))
```

```
## Warning: Scaling data with a low number of groups may produce misleading
## results
```

```r
  p
```


Identity
EO_enrich
LO_enrich
Common_clus
RRM2
TUBA1B
STMN1
TYMS
TOP2A
KLRC1
FOS
ZNF331
TNFAIP3
BTG2
FCRL3
KLRG1
TNFSF4
INPP1
TTN
0.5
0.0
−0.5
−1.0
Percent Expressed
0
25
50
75
100