

Data availability

```
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109555
```

load TPM matrix into R studio

```
test<-
read.table( "/Users/keqinliu/studying_document/bioinfo/genes_involved_in_embryo_irreversible_development_analysis_project/human_embryo_implantation_rawdata/GSE109555_All_Embryo TPM.txt" )
```

Logarithmic processing of matrices

```
test <- log10(test + 1)
```

sort by embryo stage(day)

```
library(dplyr)
only_d6 <- select(test, starts_with("D6"))
only_d8 <- select(test, starts_with("D8"))
only_d10 <- select(test, starts_with("D10"))
only_d12 <- select(test, starts_with("D12"))
only_d14 <- select(test, starts_with("D14"))

all_stage <- cbind(only_d6,only_d8,only_d10,only_d12,only_d14)
```

Create Seurat Object

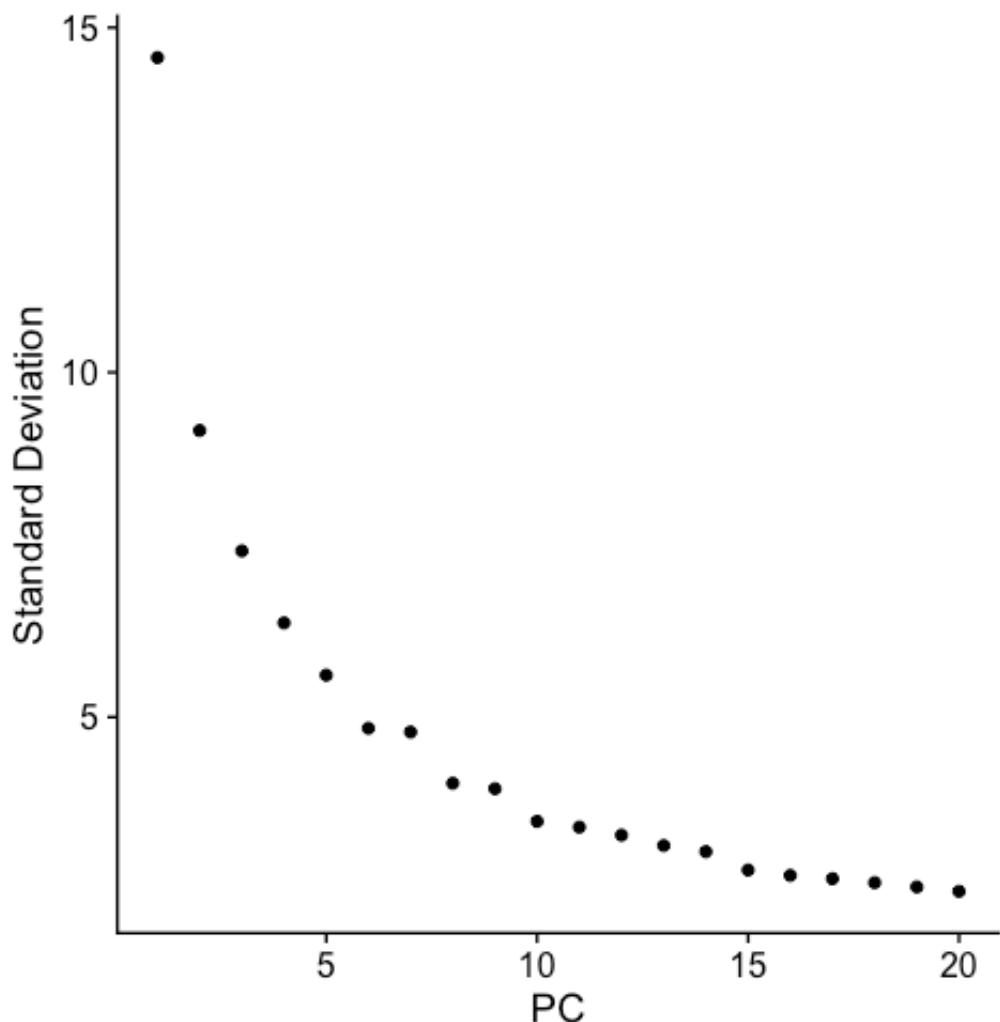
```
library(Seurat)
all_stage_new <- CreateSeuratObject(
  counts = all_stage, # test is all sample_gene matrix
  project = "SeuratProject",
  assay = "RNA",
  min.cells = 0,
  min.features = 0,
  names.field = 1,
  names.delim = "-",
  meta.data = NULL
)
```

Find Variable Features

```
all_stage_new_gene <- FindVariableFeatures(all_stage_new, selection.method =  
"vst", nfeatures = 2000)  
all_stage_new_top10 <- head(VariableFeatures(all_stage_new_gene), 10)  
all_stage_new_genes <- rownames(all_stage_new)  
all_stage_new <- ScaleData(all_stage_new, features = all_stage_new_genes)  
variable_all_stage_new_gene <- FindVariableFeatures(object = all_stage_new)
```

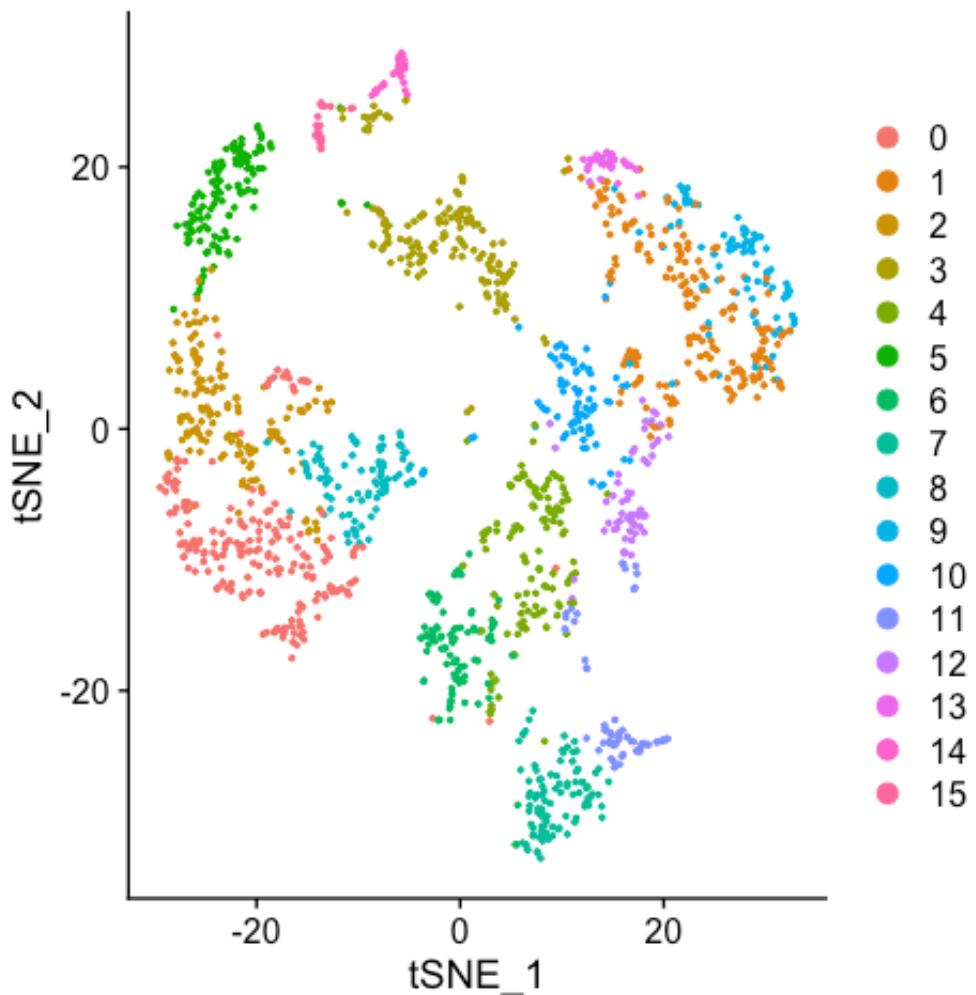
do elbowplot

```
all_stage_new <- RunPCA(all_stage_new, features = VariableFeatures(object =  
variable_all_stage_new_gene))  
ElbowPlot(all_stage_new)
```



do TSNEplot

```
all_stage_new <- FindNeighbors(all_stage_new, dims = 1:15)  
all_stage_new <- FindClusters(all_stage_new, dims = 1:15, print = FALSE)  
all_stage_new <- RunTSNE(all_stage_new, dims.use = 1:15)  
TSNEPlot(all_stage_new, pt.size = 0.5)
```

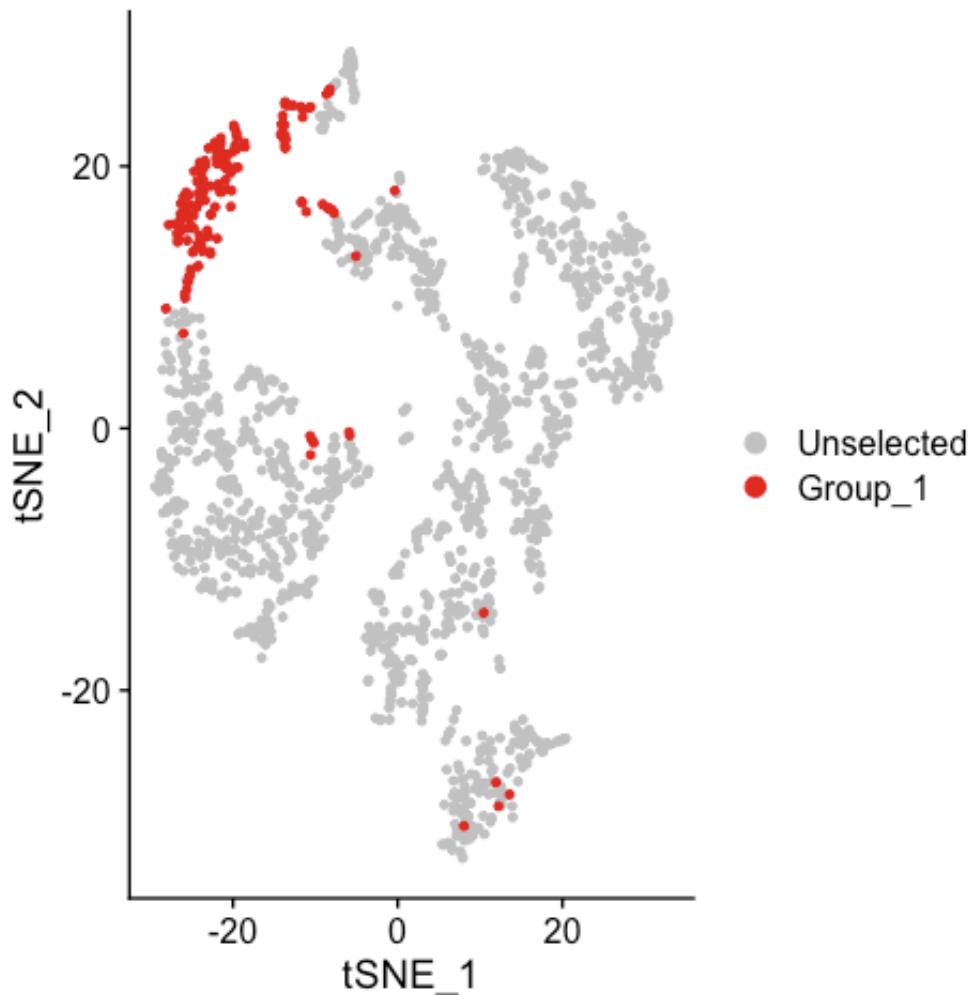


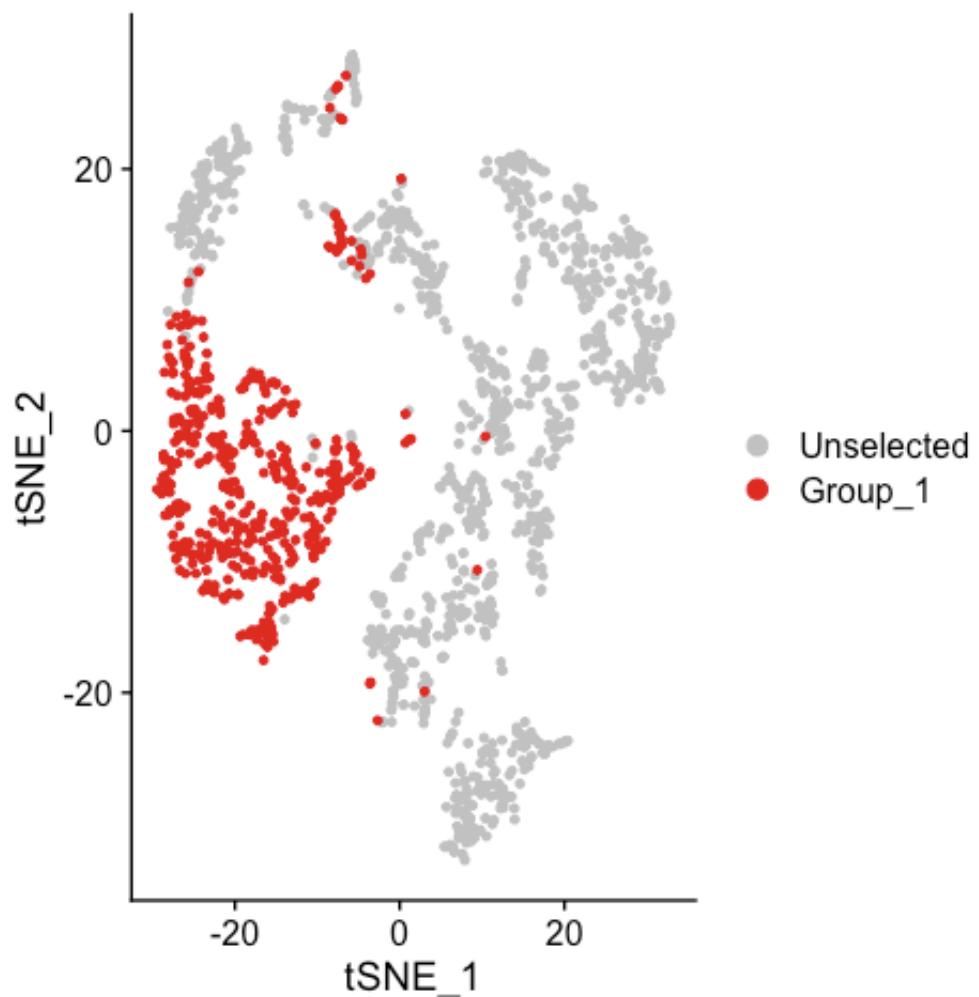
show the embryonic cells at each stage separately (day 6 to day 14) in the total cluster results

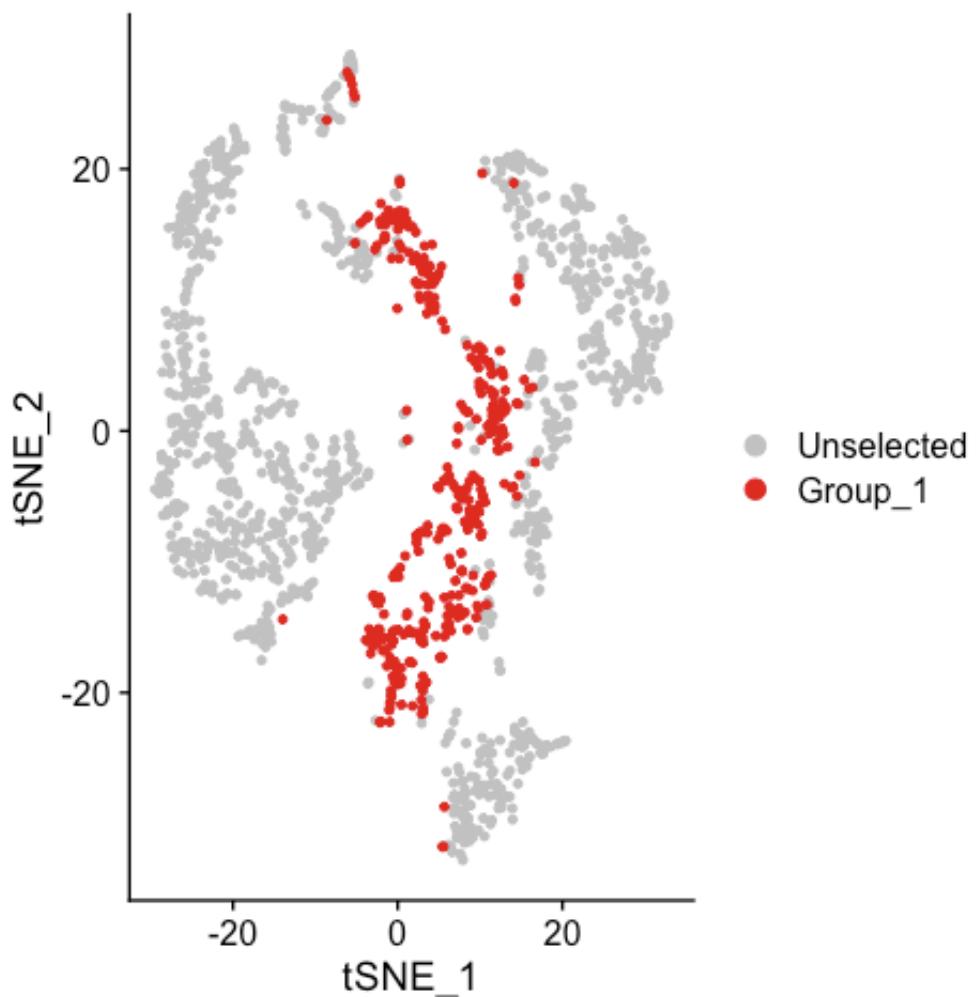
```
only_d6 <- CreateSeuratObject(
  counts = only_d6, # test is all sample_gene matrix
  project = "SeuratProject",
  assay = "RNA",
  min.cells = 0,
  min.features = 0,
  names.field = 1,
  names.delim = "-",
  meta.data = NULL
) #Create Seurat Object

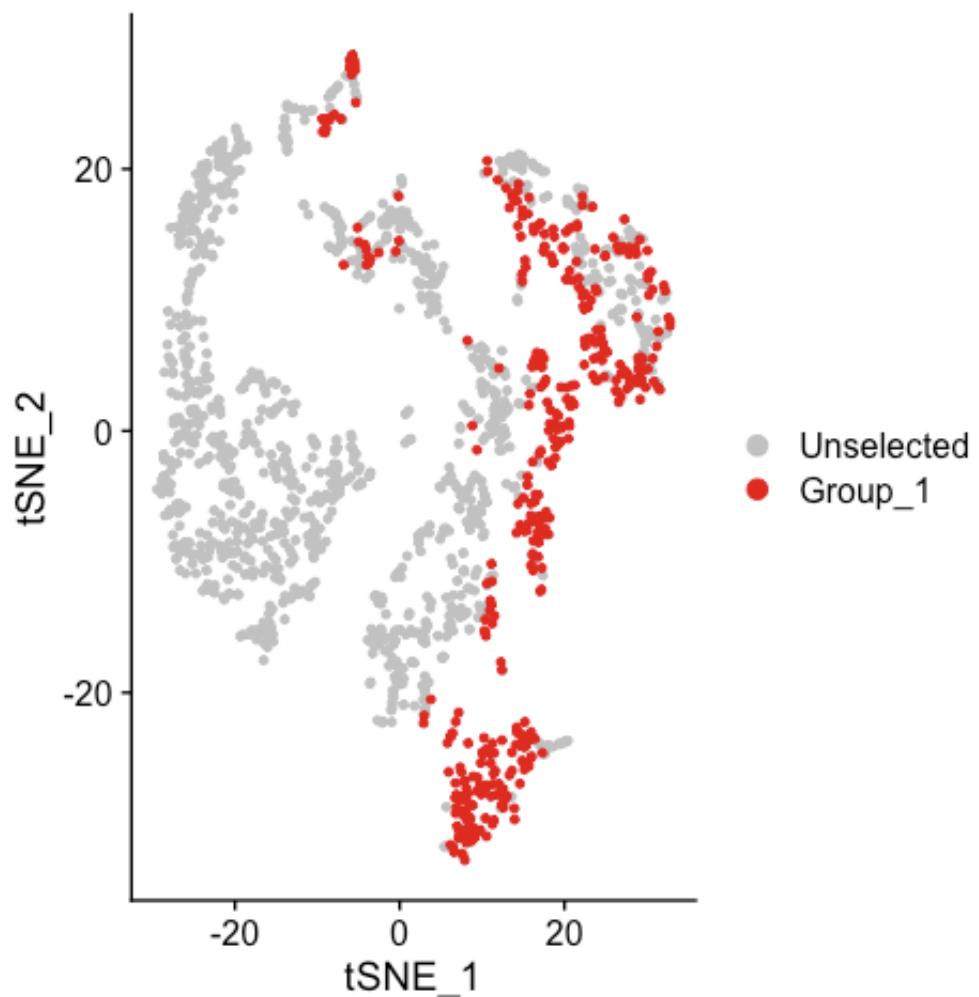
only_d6_gene <- FindVariableFeatures(only_d6, selection.method = "vst",
nfeatures = 2000)
only_d6_gene <- rownames(only_d6)
only_d6 <- ScaleData(only_d6, features = only_d6_gene)
variable_only_d6_gene <- FindVariableFeatures(object = only_d6)
only_d6 <- RunPCA(only_d6, features = VariableFeatures(object =
variable_only_d6_gene))
only_d6 <- JackStraw(only_d6, num.replicate = 100)
only_d6 <- ScoreJackStraw(only_d6, dims = 1:20)
```

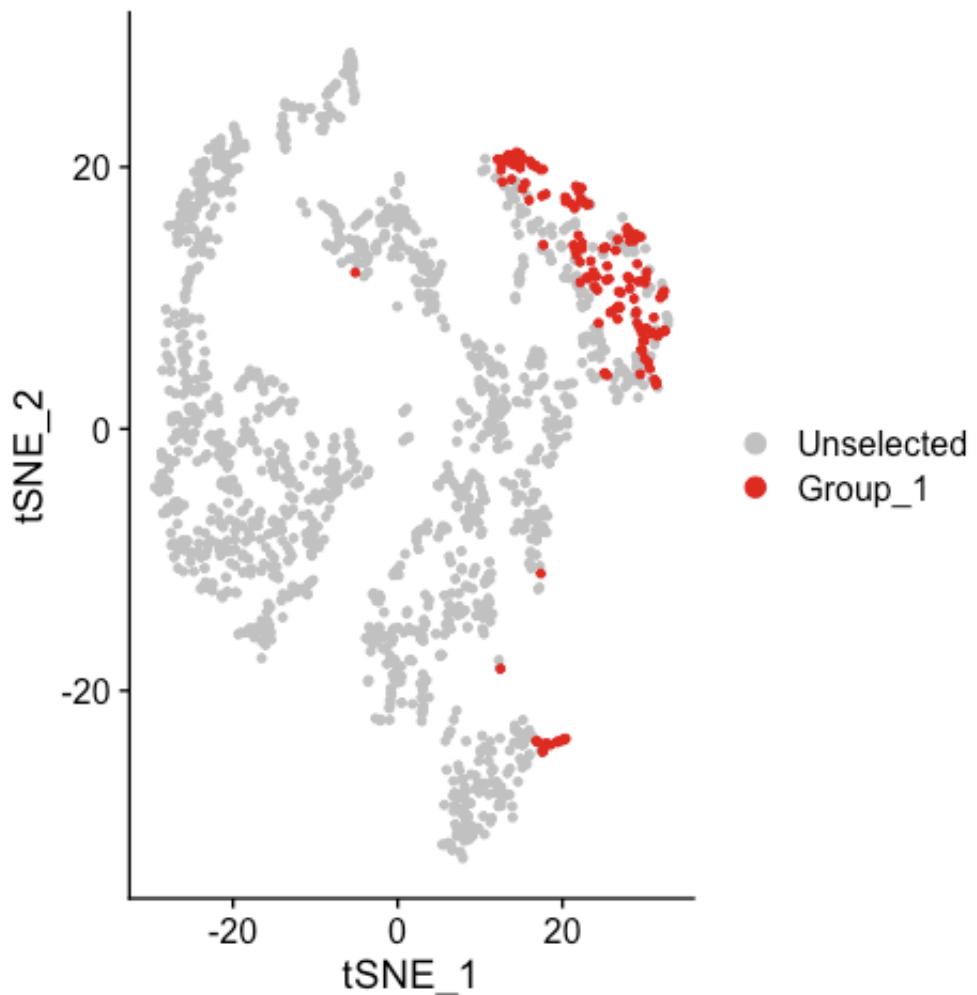
```
only_d6 <- FindNeighbors(only_d6, dims = 1:10)
only_d6 <- FindClusters(only_d6, resolution = 0.5)
only_d6 <- RunUMAP(only_d6, dims = 1:10)
```









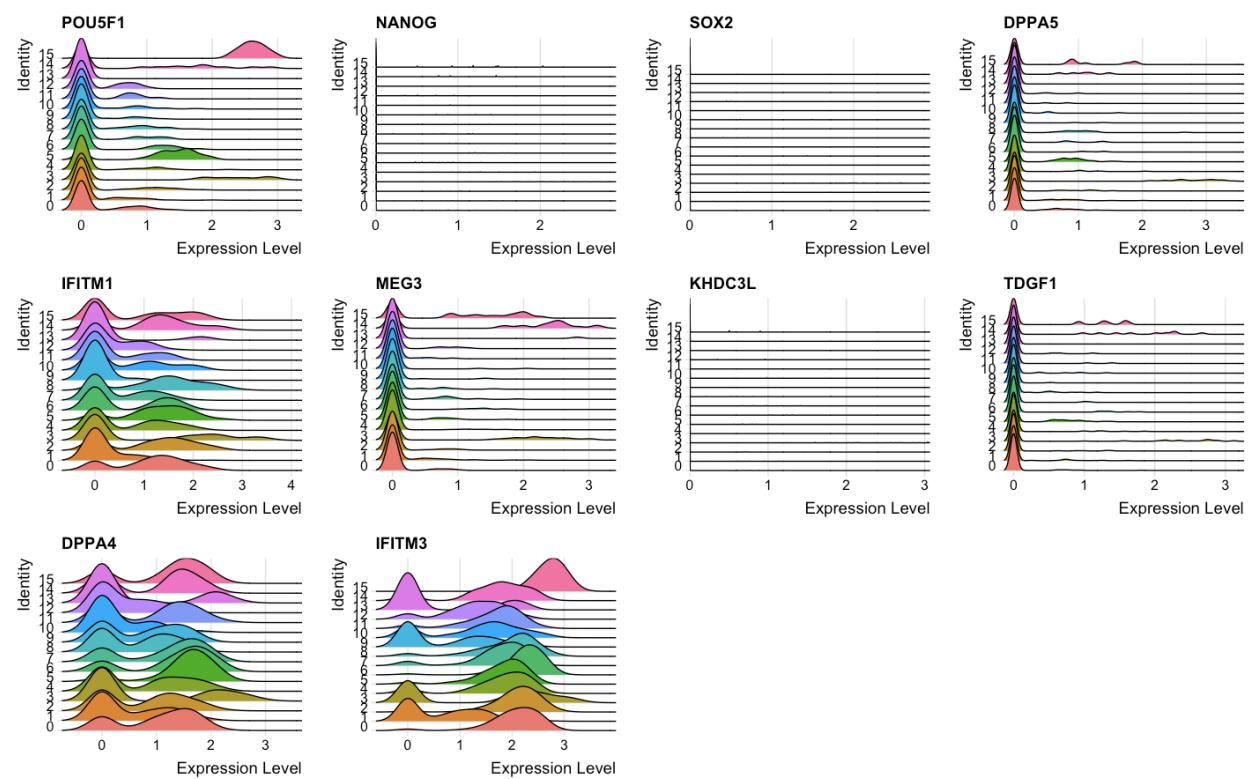
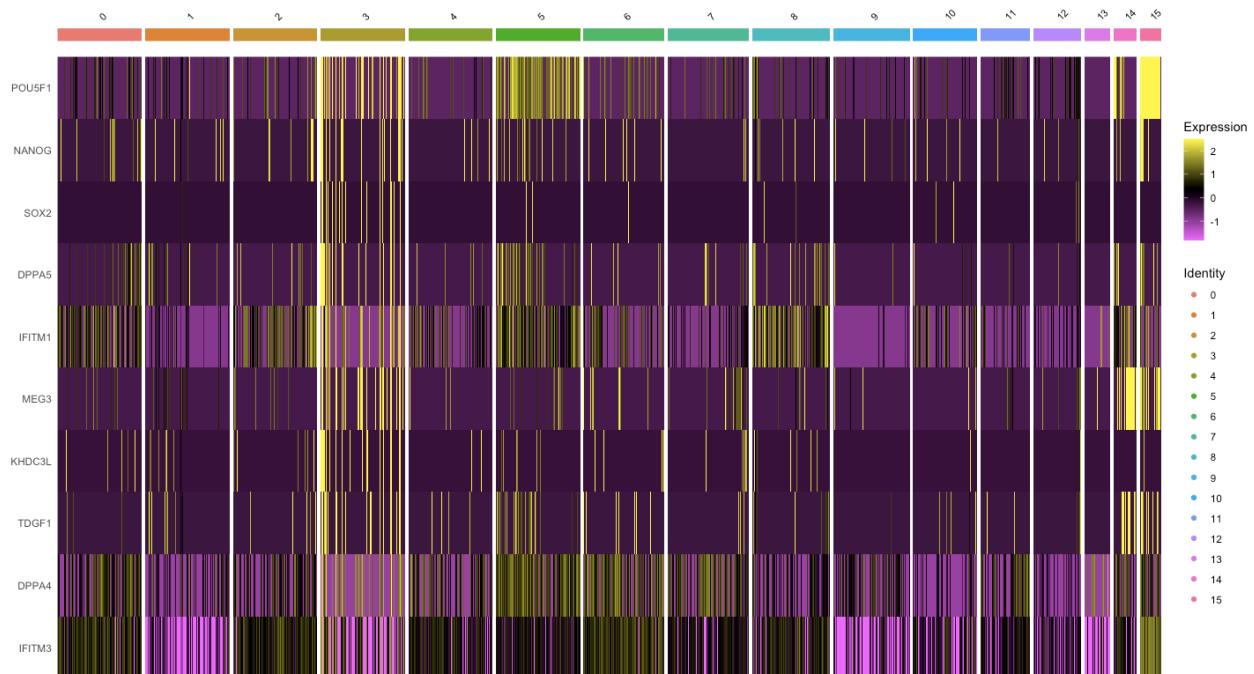


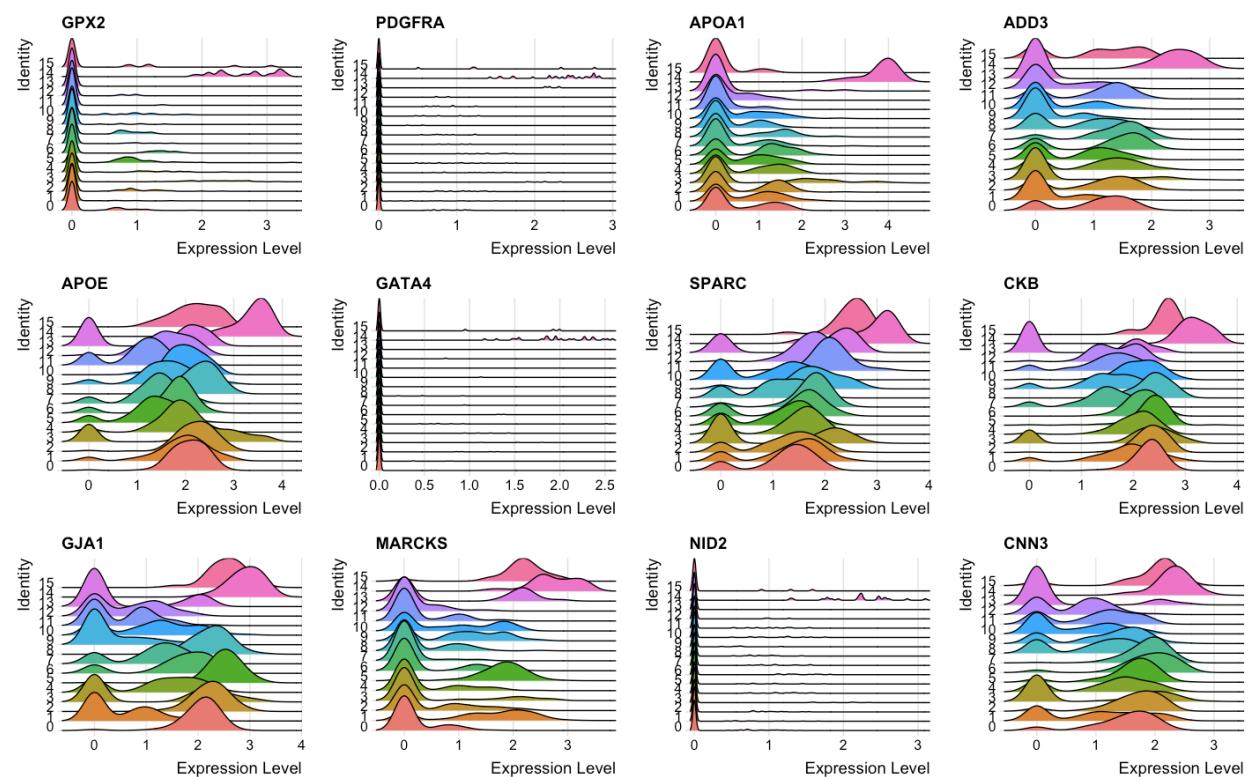
then do heatmap and ridgeplot

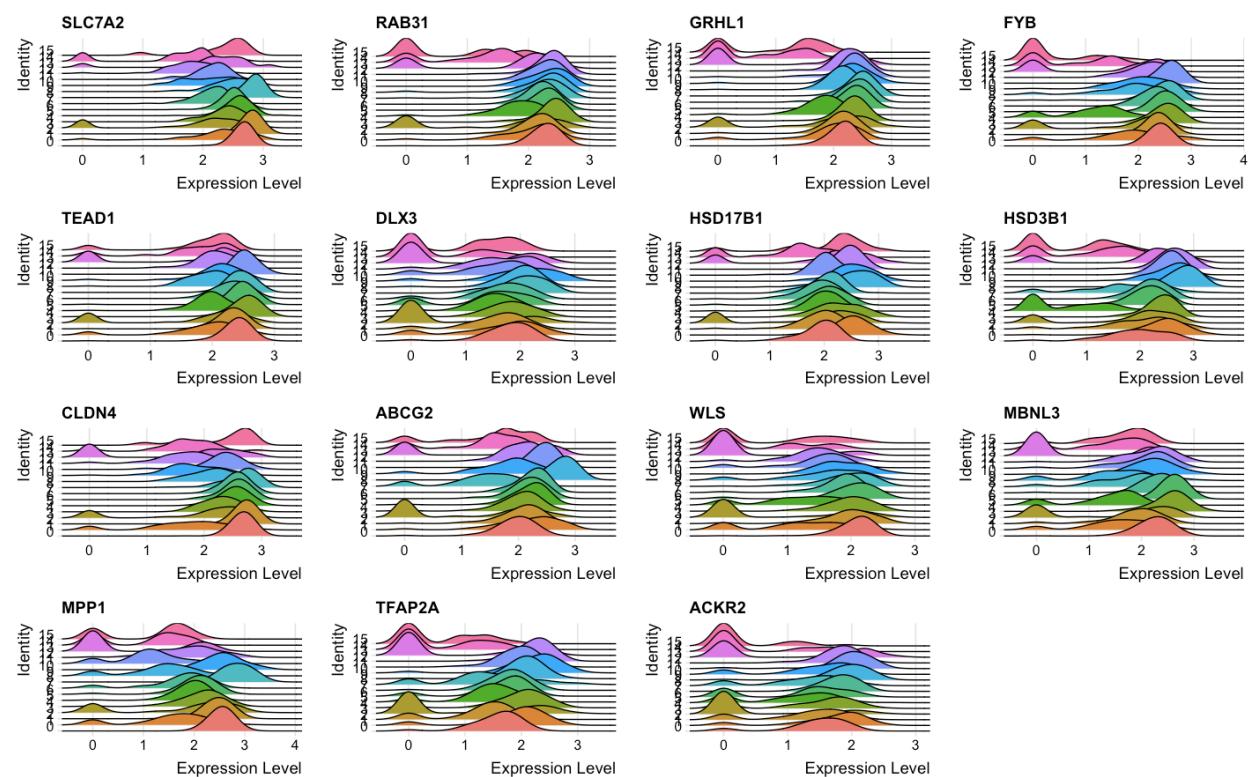
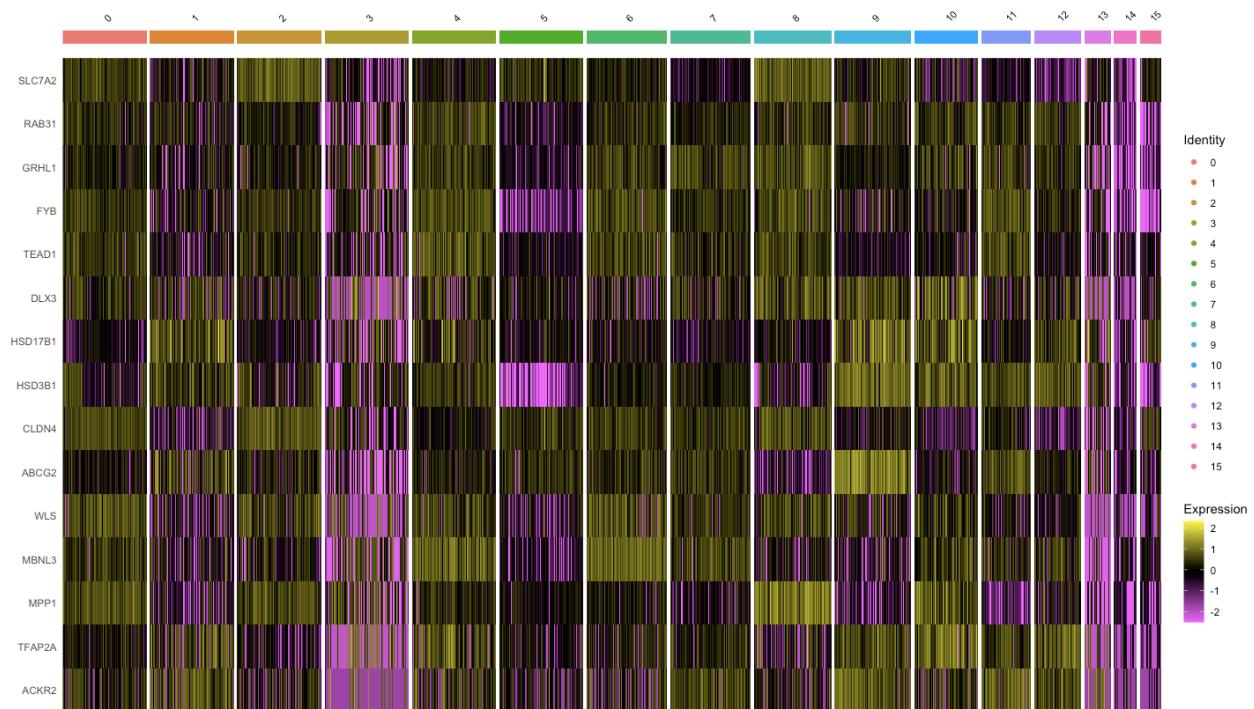
```

epi_feature <- c("POU5F1", "NANOG", "SOX2", "DPPA5", "IFITM1", "MEG3",
"KHDC3L", "TDGF1", "DPPA4", "IFITM3")
te_feature <- c("SLC7A2", "RAB31", "GRHL1", "FYB", "TEAD1", "DLX3", "HSD17B1",
"HSD3B1", "CLDN4", "ABCG2", "WLS", "MBNL3", "MPP1", "TFAP2A", "ACKR2")
pe_feature <- c("GPX2", "PDGFRA", "APOA1", "ADD3", "APOE", "GATA4", "SPARC",
"CKB", "GJA1", "MARCKS", "NID2", "CNN3")
icm_feature <- c("NANOG", "GATA6", "EOMES", "SOX2", "POU5F1", "FGF4")
RidgePlot(all_stage_new, features = epi_feature)
RidgePlot(all_stage_new, features = te_feature)
RidgePlot(all_stage_new, features = pe_feature)
DoHeatmap(subset(all_stage_new, downsample = 100), features = epi_feature, size
= 3)
DoHeatmap(subset(all_stage_new, downsample = 100), features = te_feature, size
= 3)
DoHeatmap(subset(all_stage_new, downsample = 100), features = pe_feature, size
= 3)

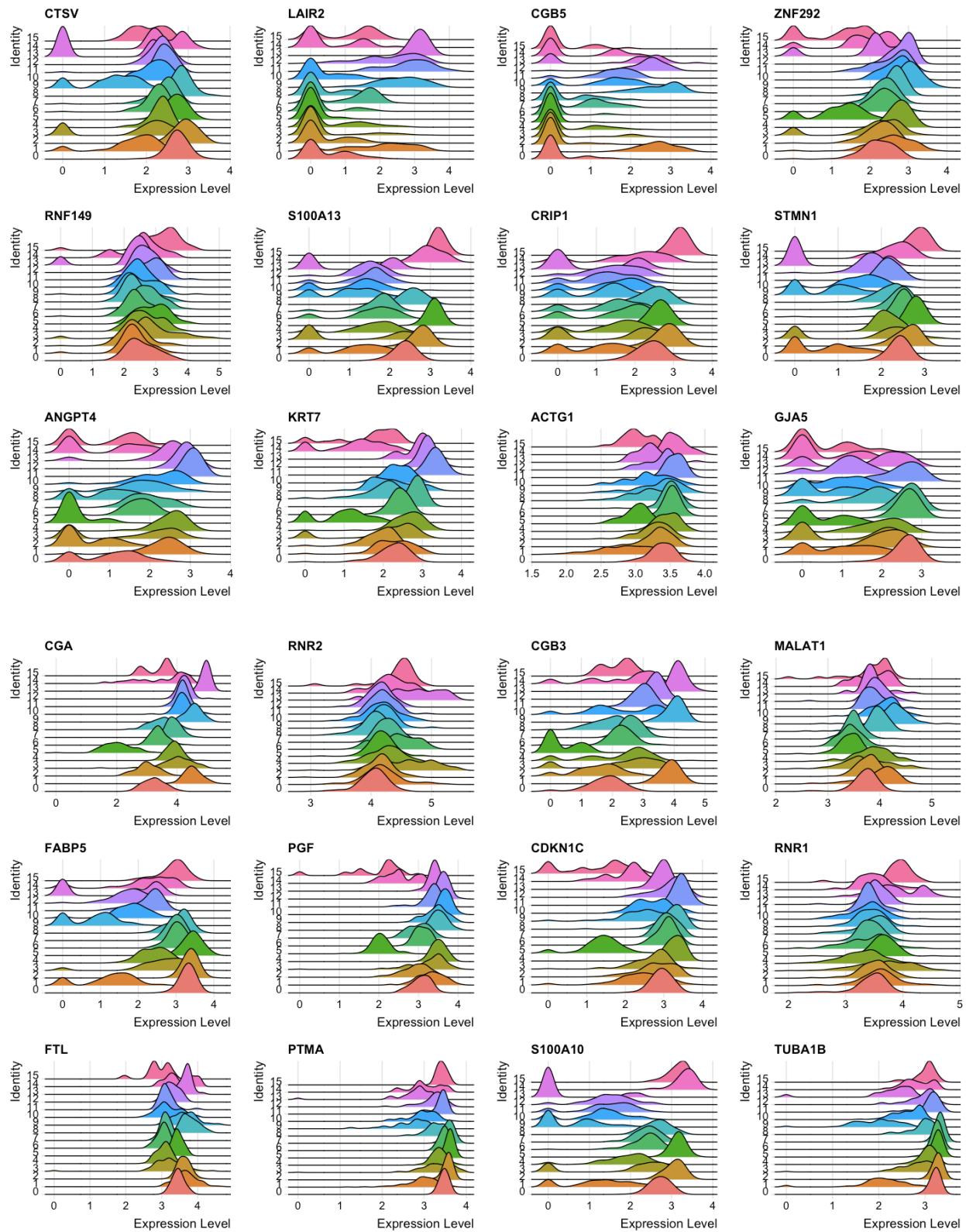
```

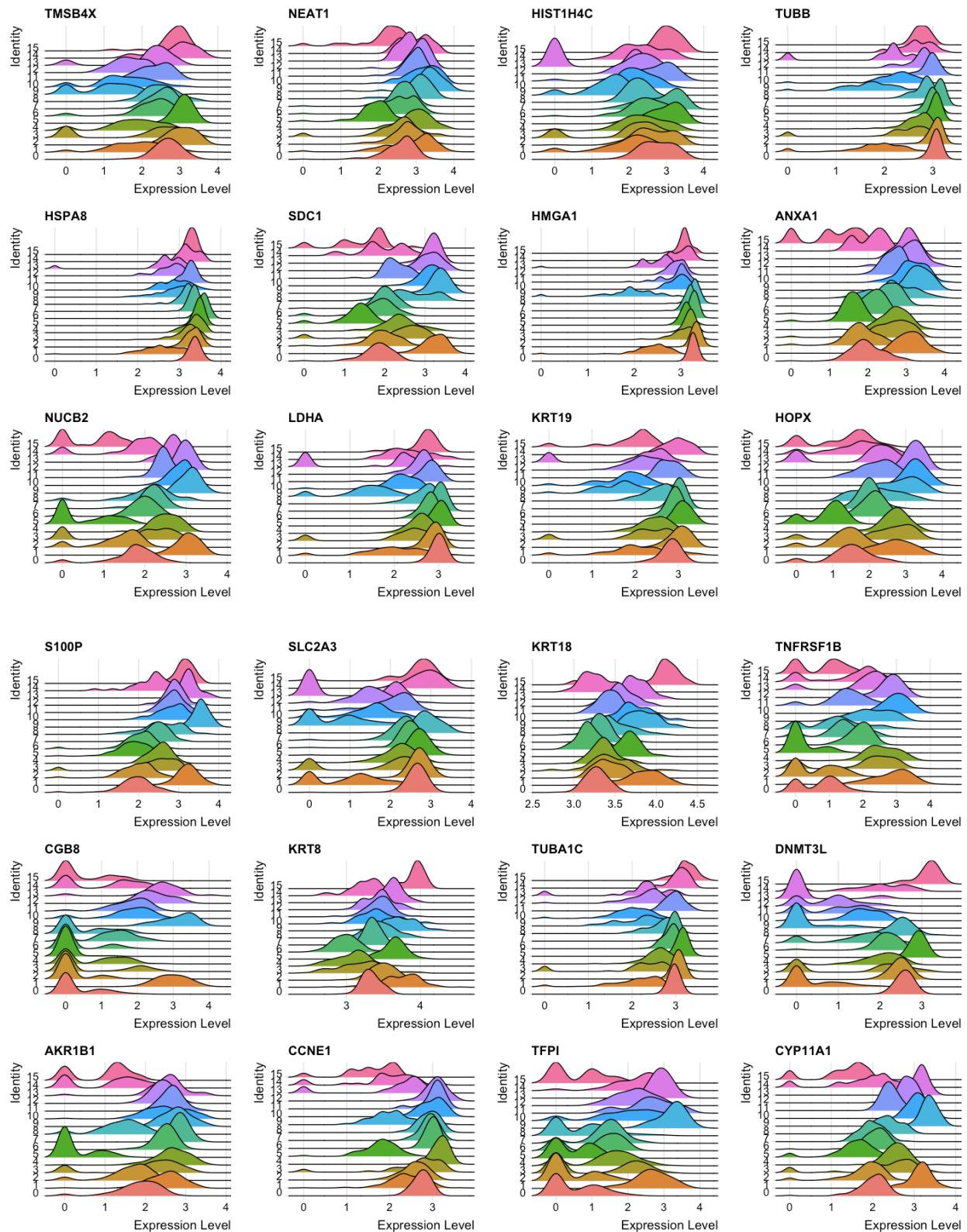


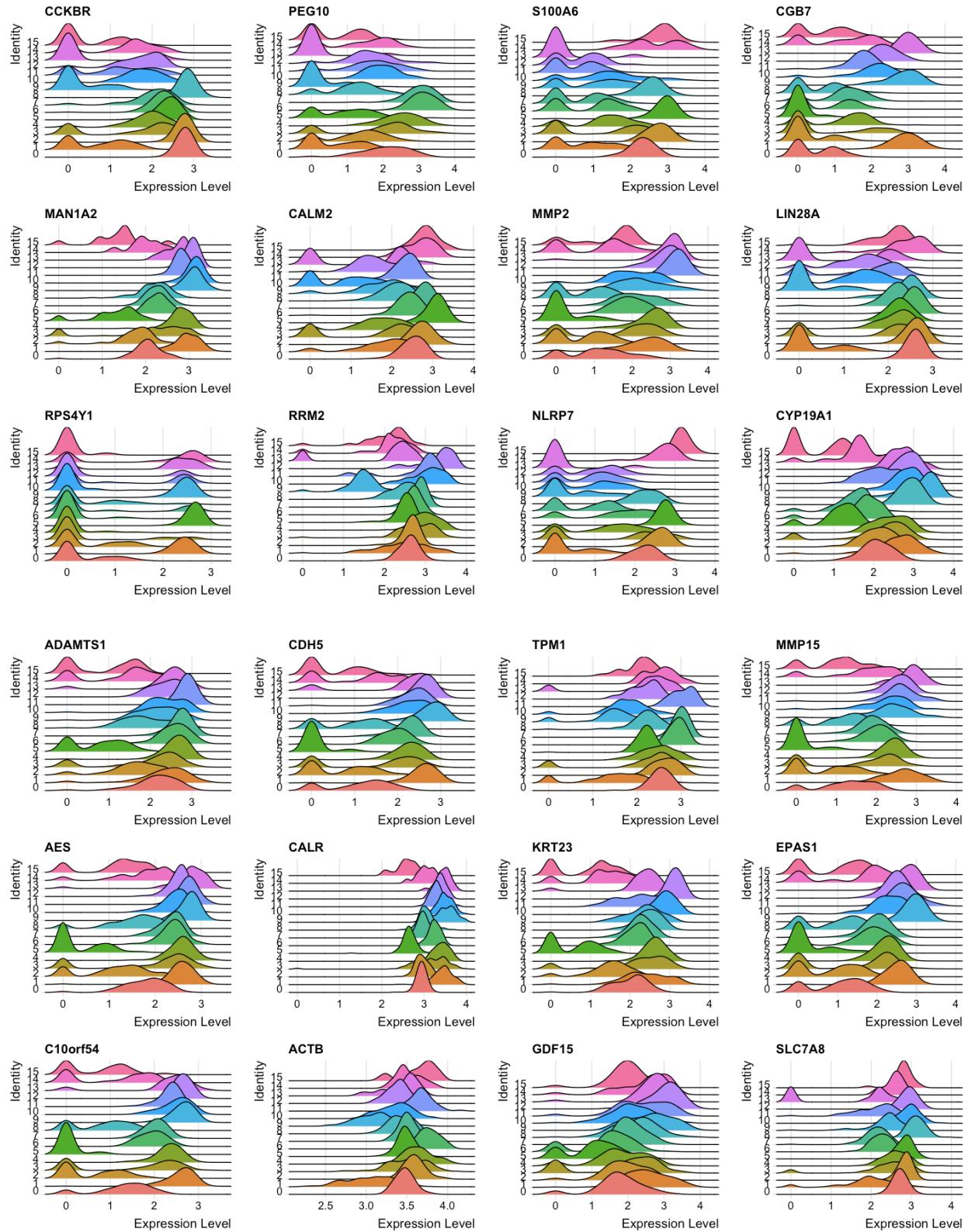


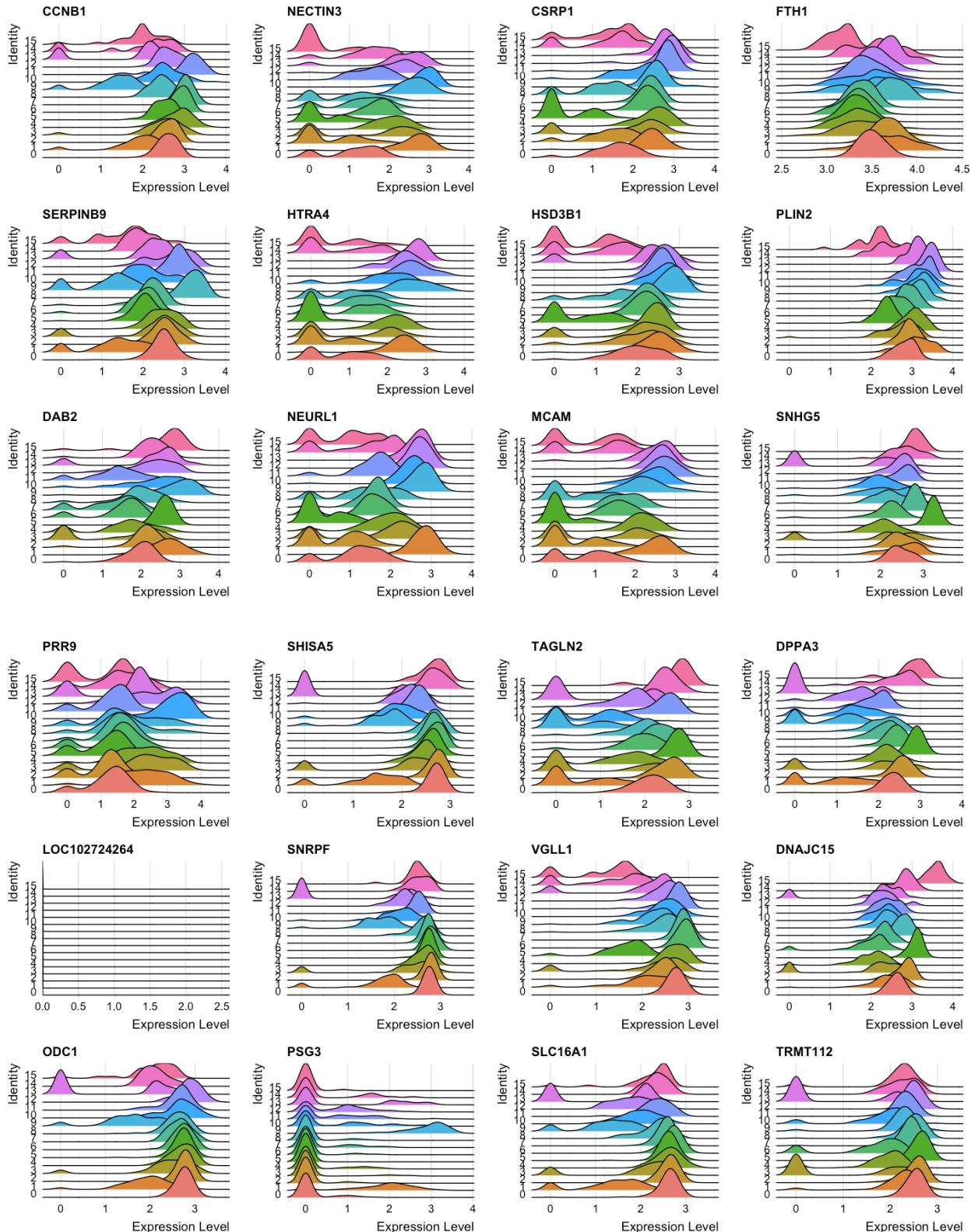


Show top 100 (actually top 96) differentially expressed genes in ascending order









Show cluster results by embryonic stage in total cluster results

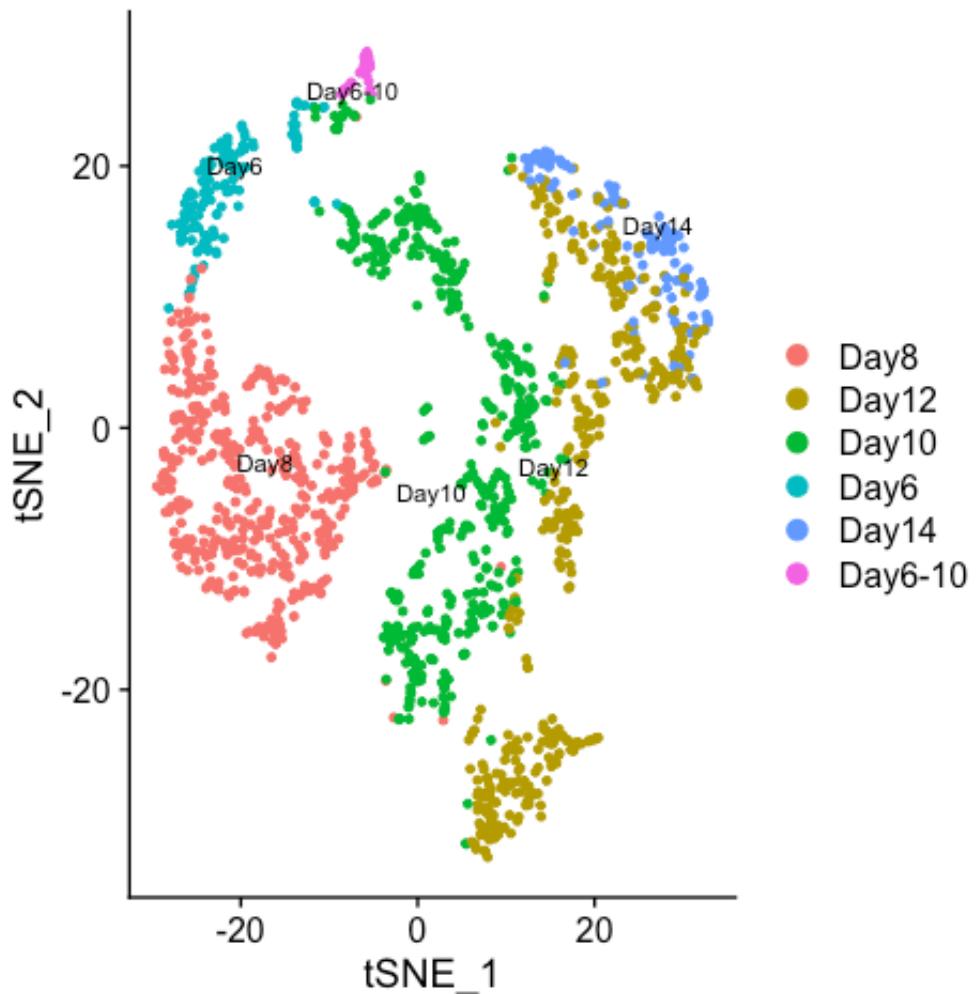
```
cell_type <- RenameIdents(object = all_stage_new,
                           "0" = "Day8",
                           "1" = "Day12",
                           "2" = "Day8",
                           "3" = "Day10",
                           "4" = "Day10",
```

```

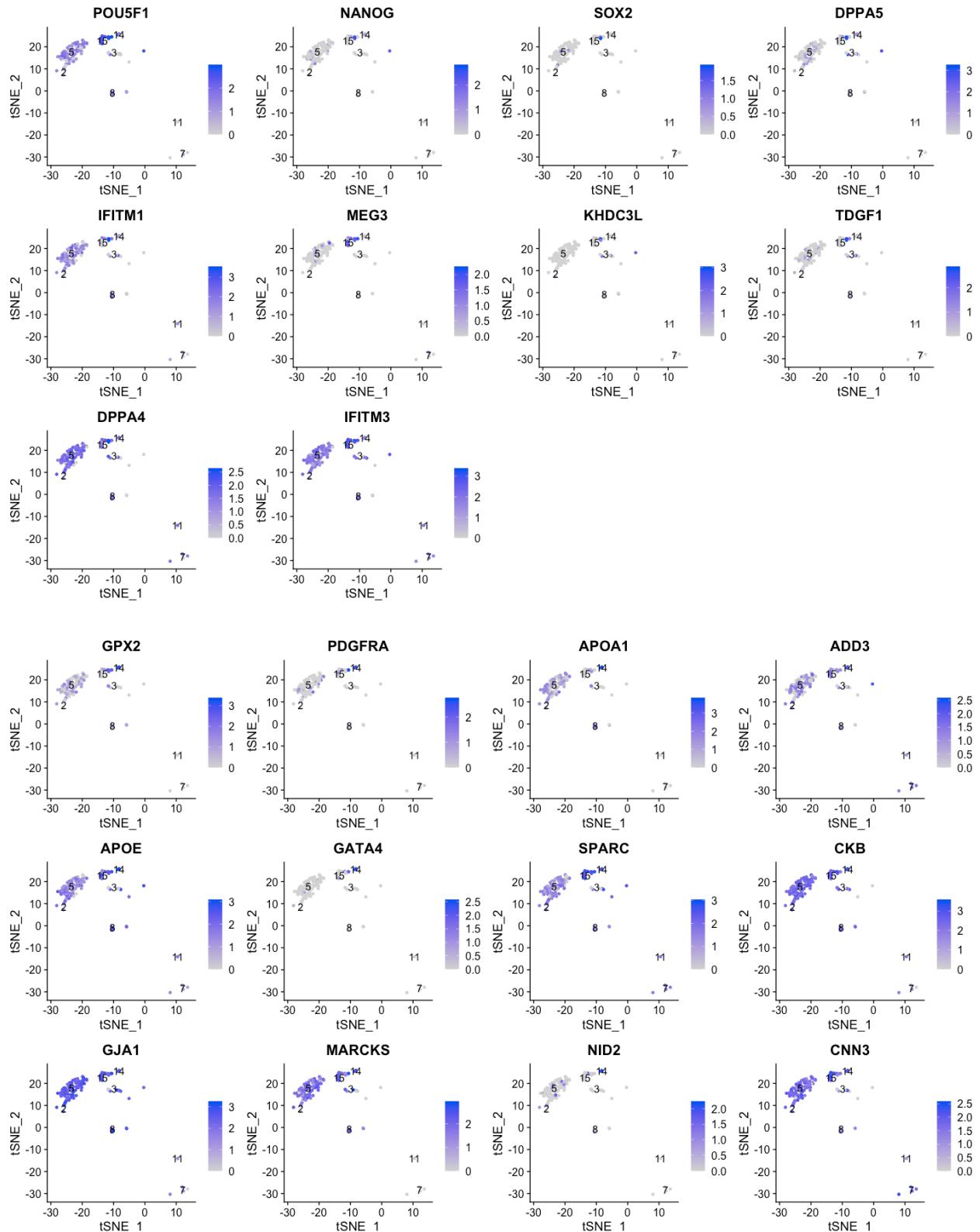
"5" = "Day6",
"6" = "Day10",
"7" = "Day12",
"8" = "Day8",
"9" = "Day14",
"10" = "Day10",
"11" = "Day12",
"12" = "Day12",
"13" = "Day14",
"14" = "Day6-10",
"15" = "Day6")

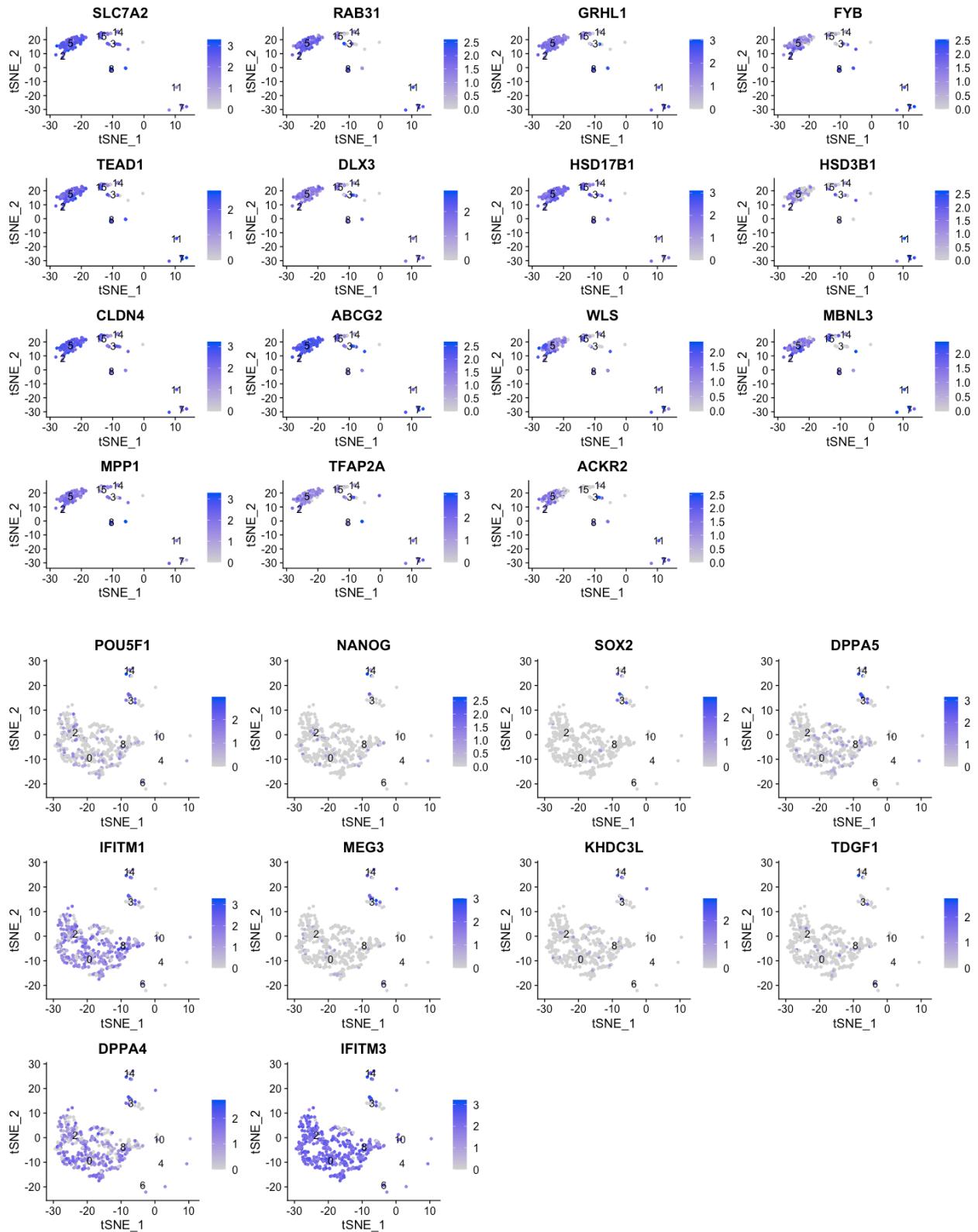
DimPlot(object = cell_type,
        reduction = "tsne",
        label = TRUE,
        label.size = 3,
        repel = TRUE)

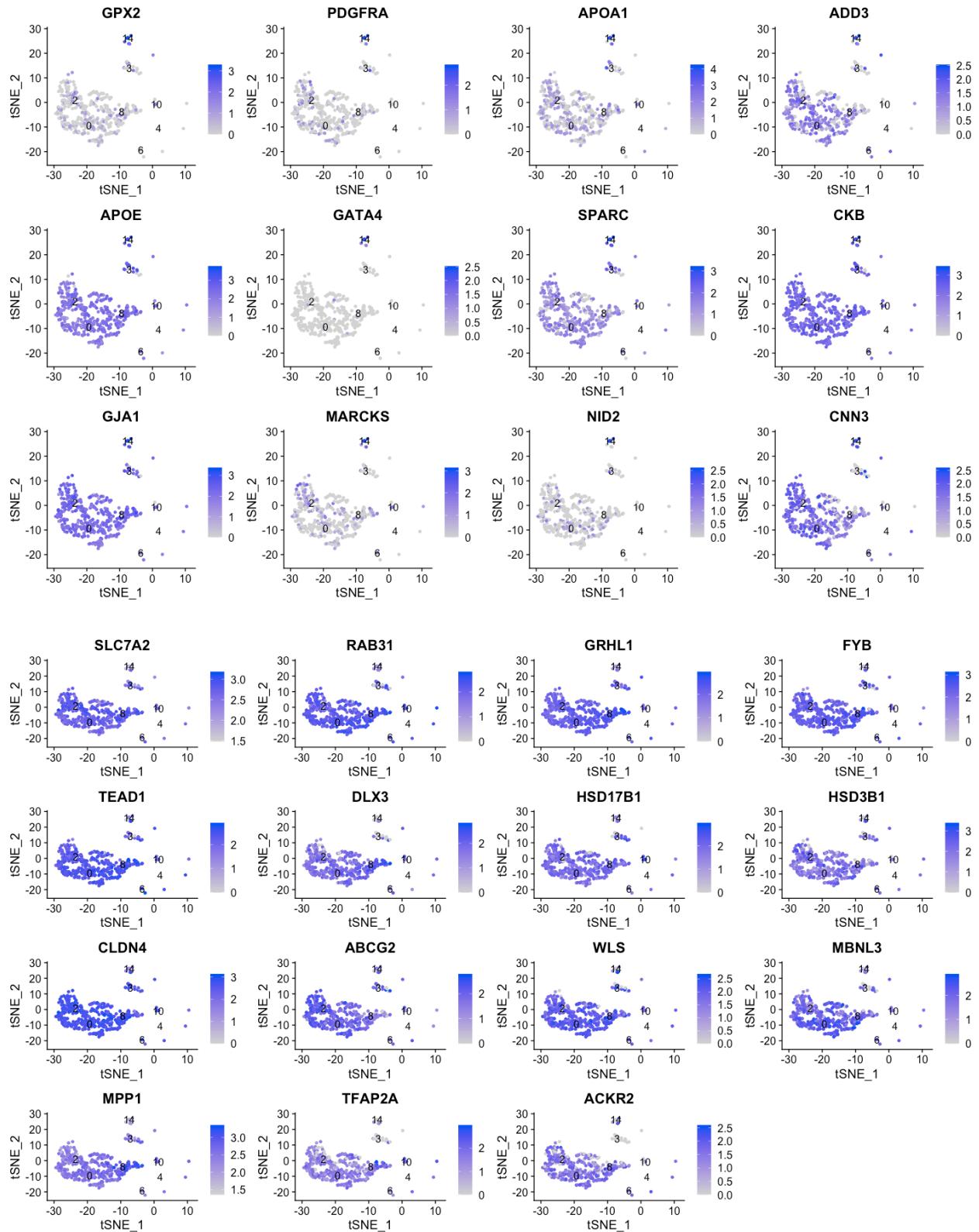
```

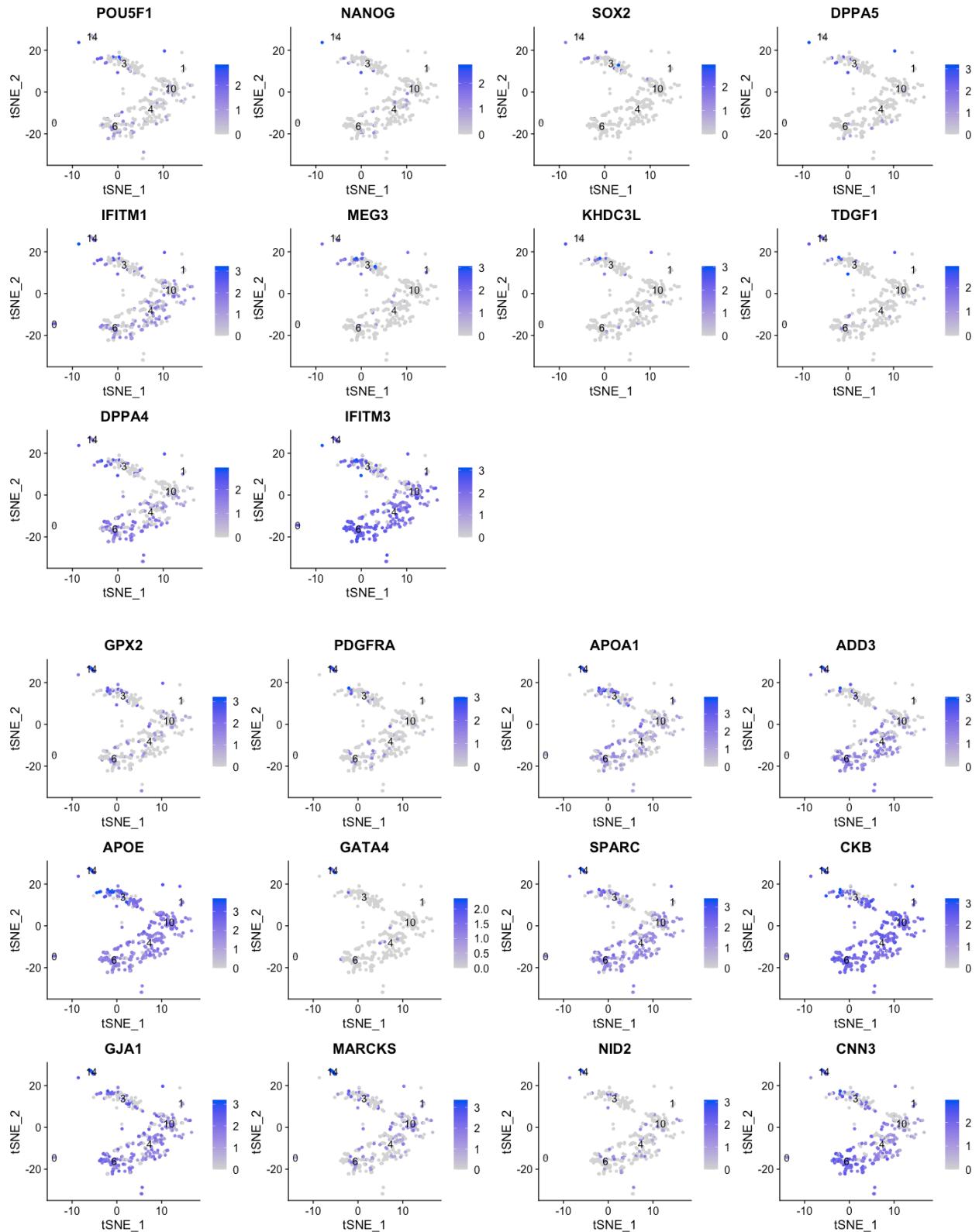


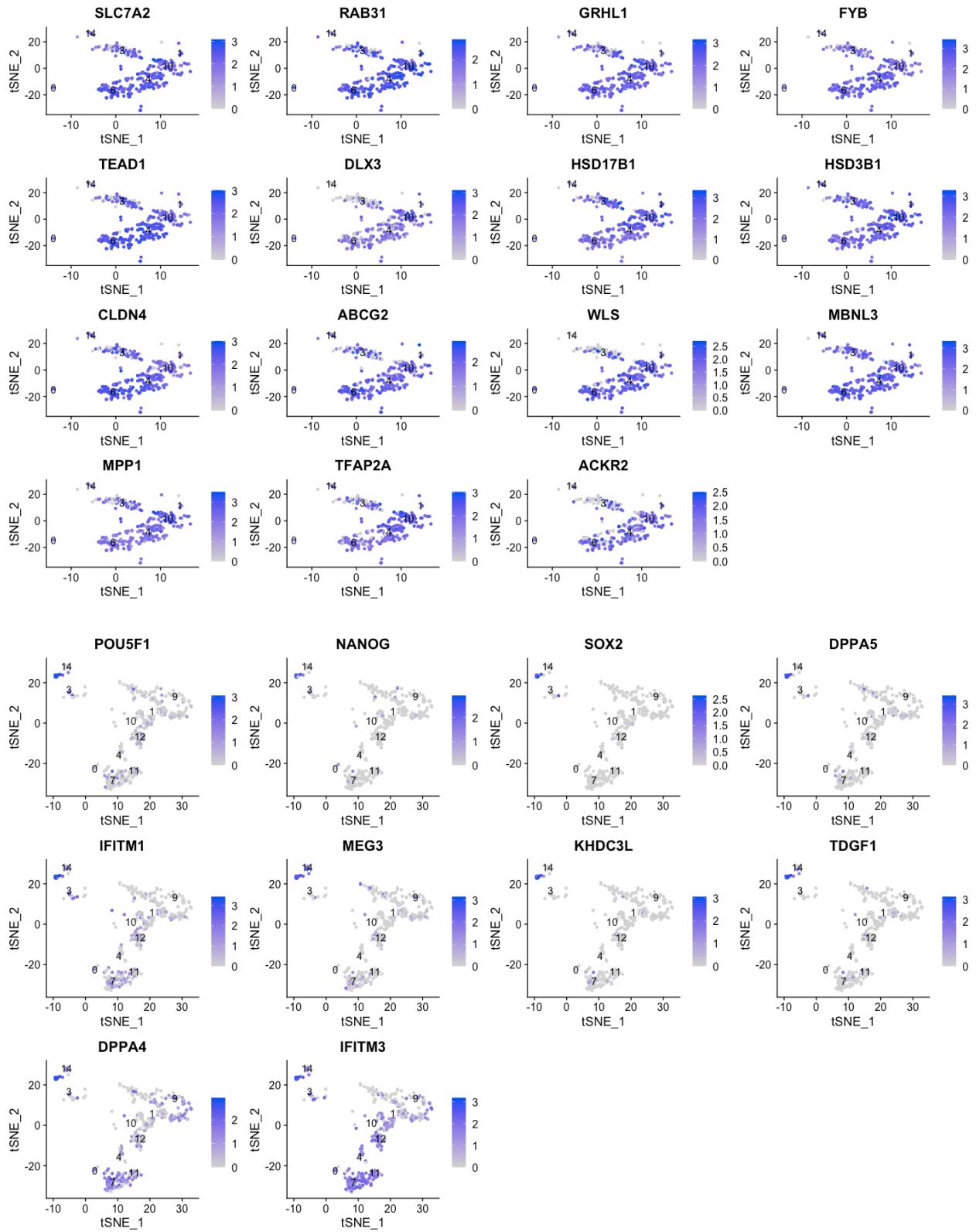
In addition, we attempted to view the labeling results of EPI, PE and TE cells at each stage of the cell mass (from day 6 to day 14).

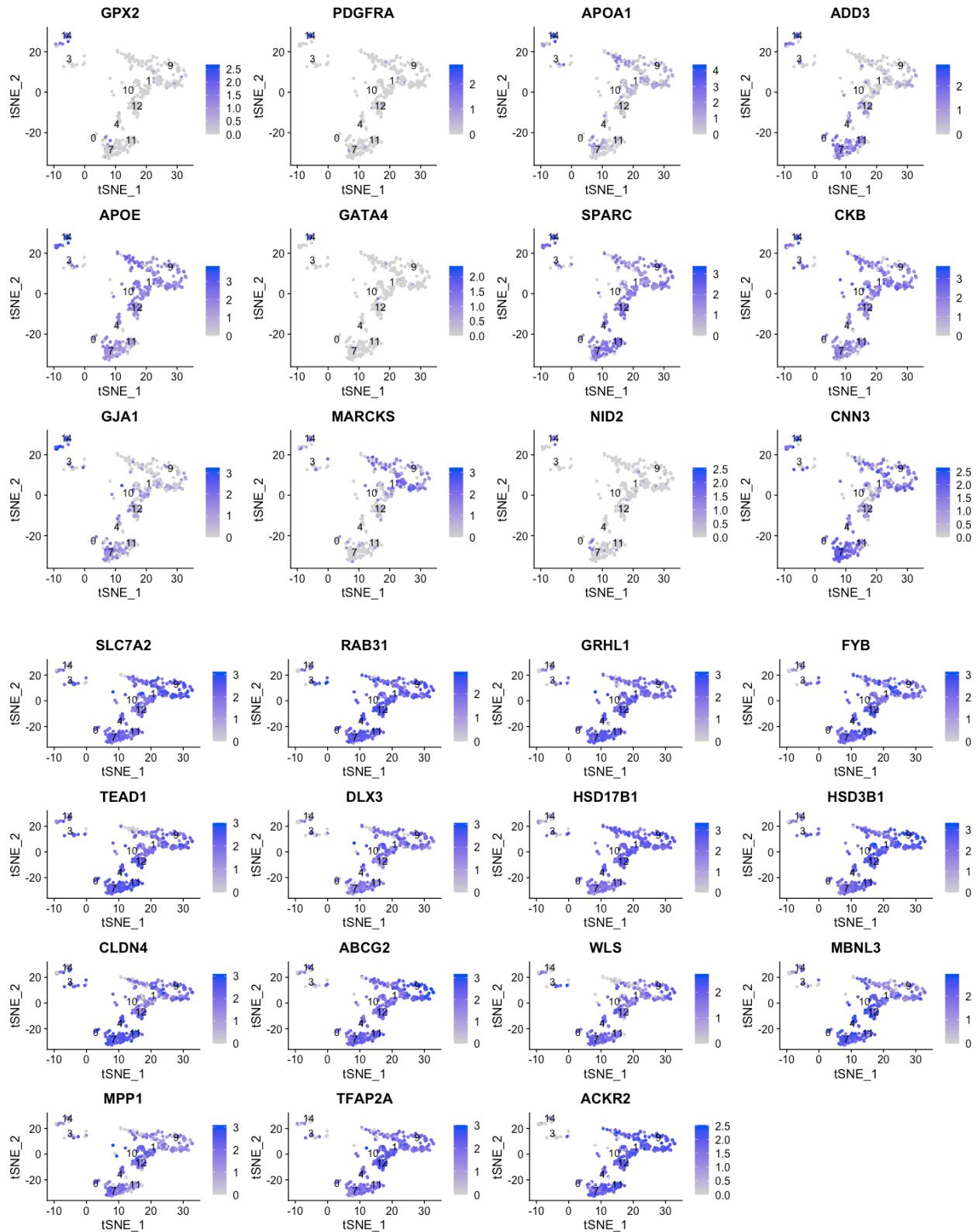


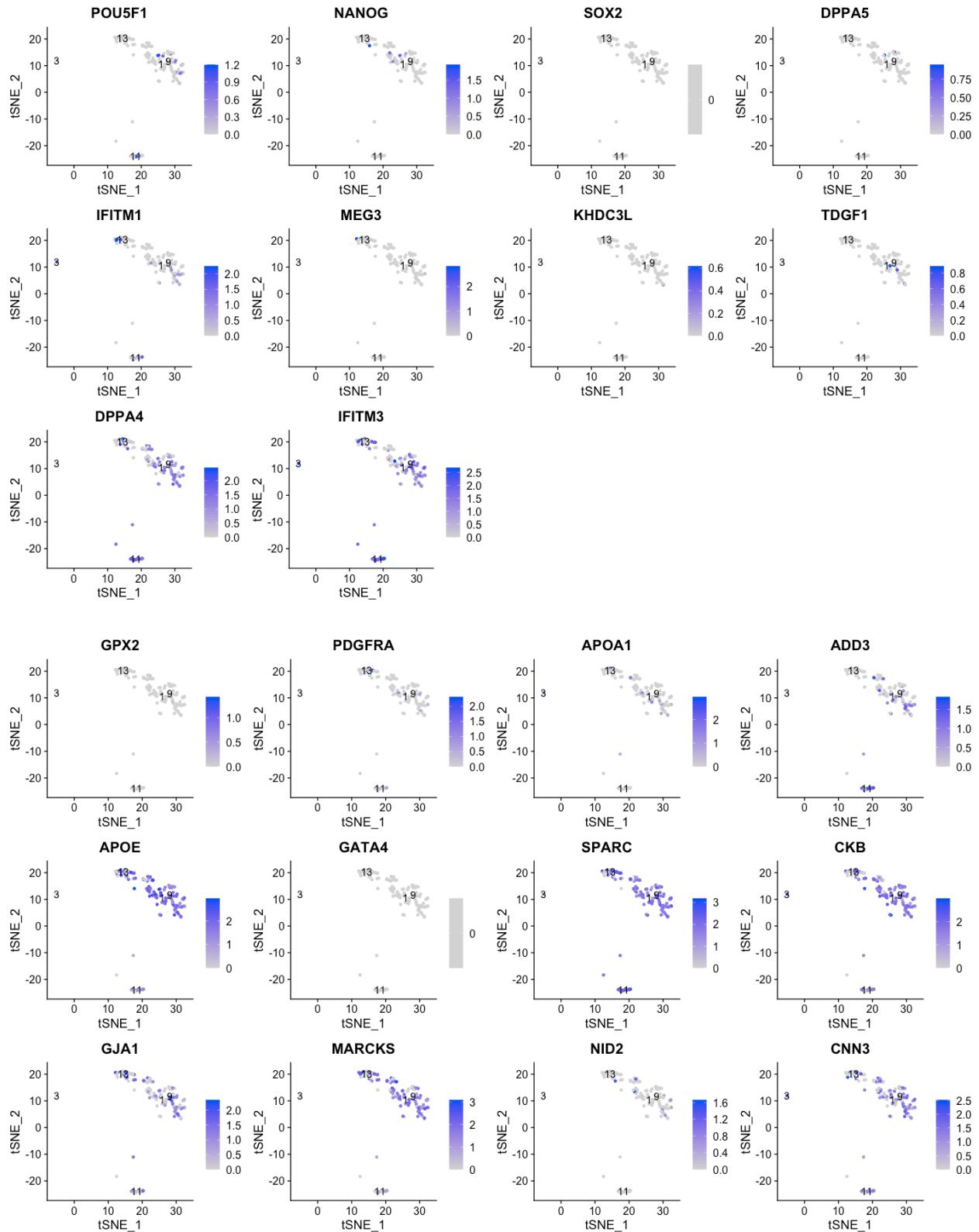


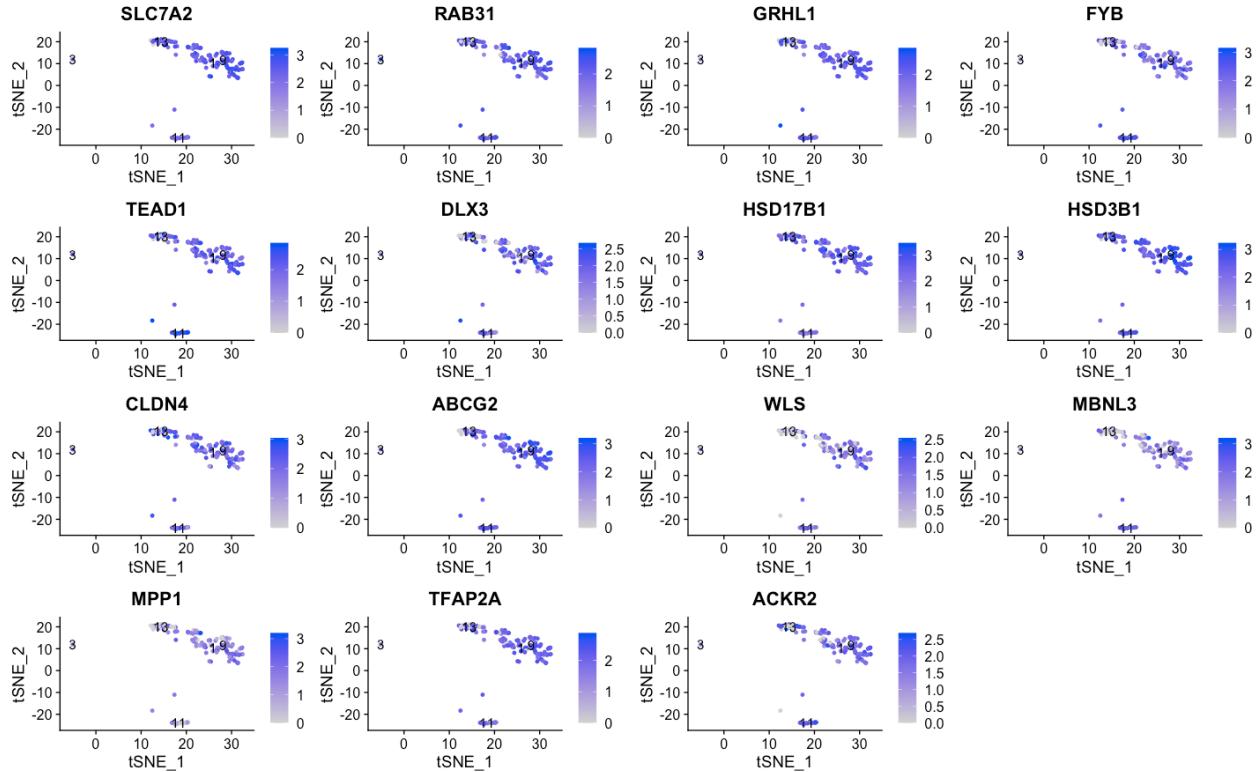








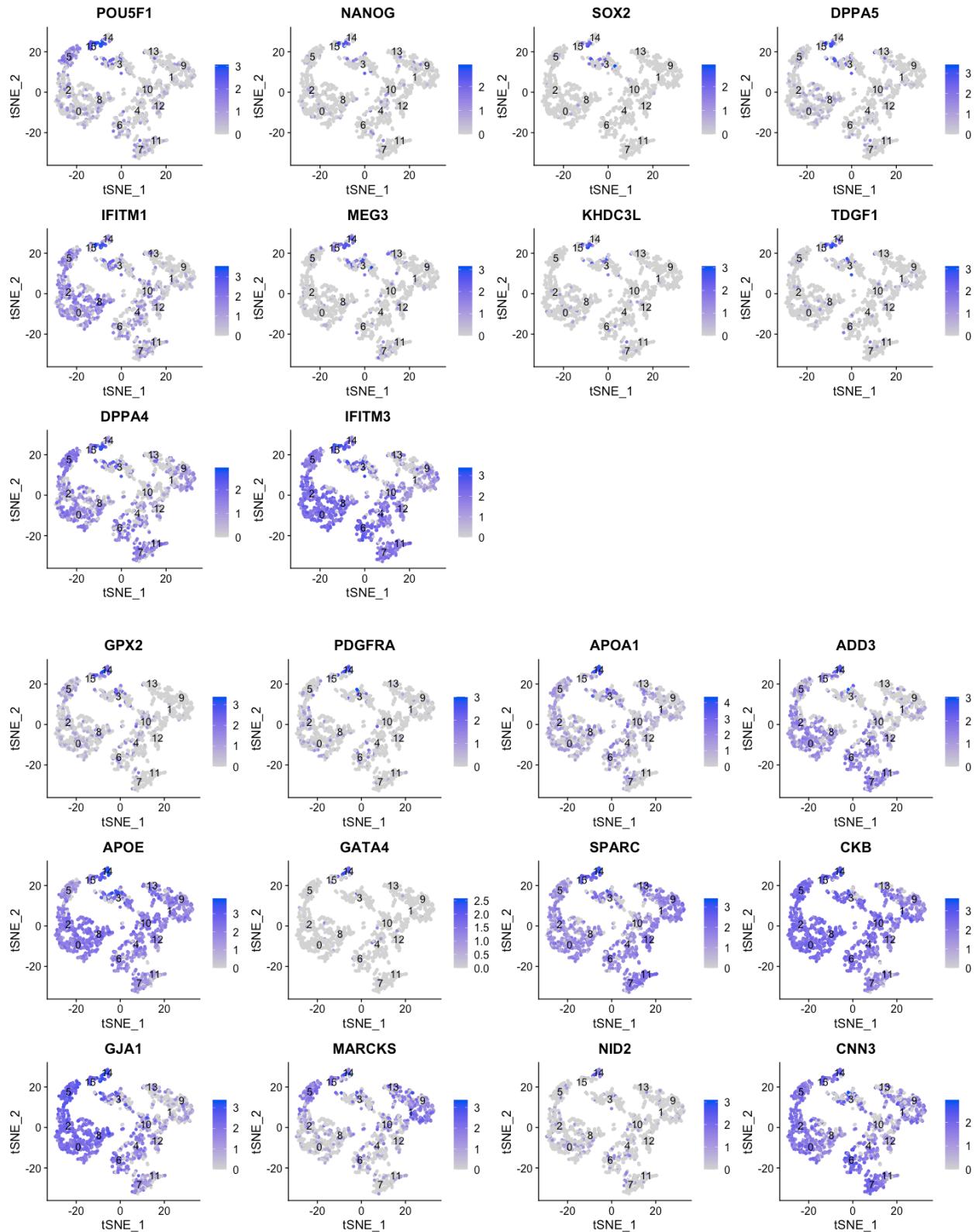


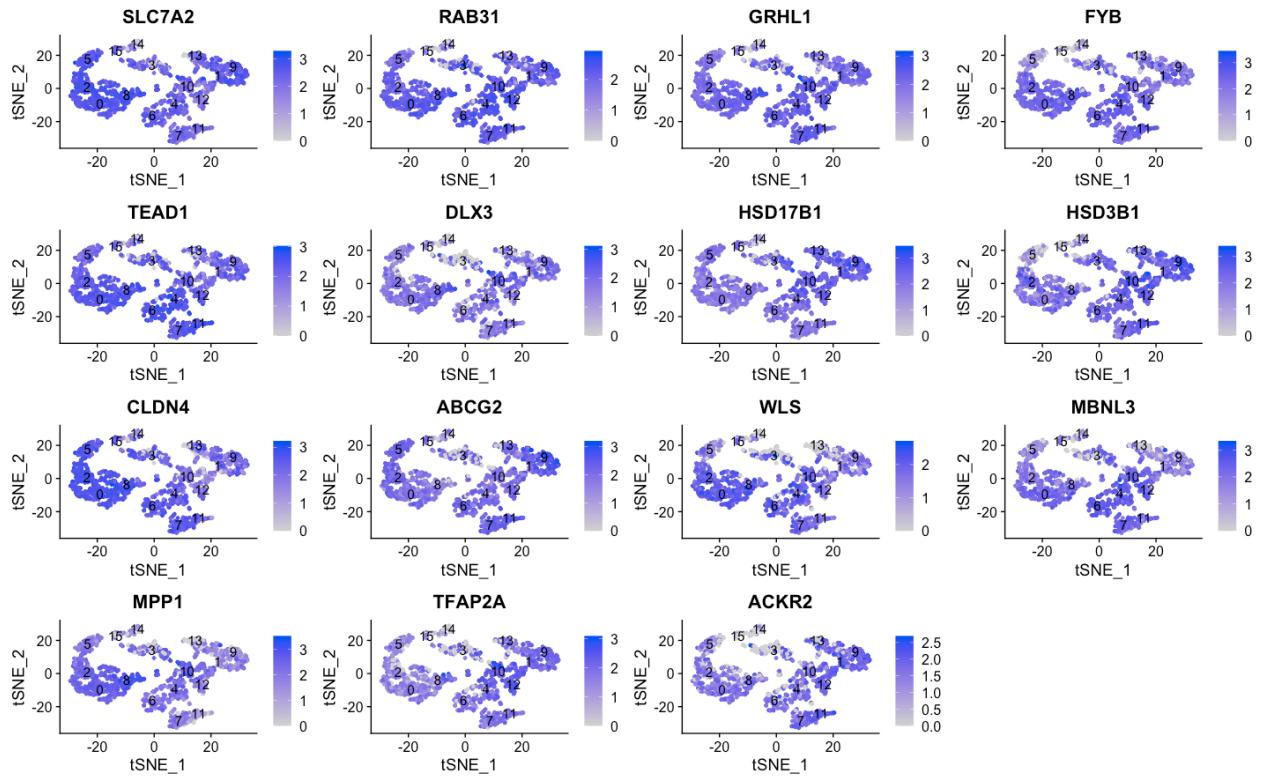


Similarly, show cluster results in the total cluster results by cell type

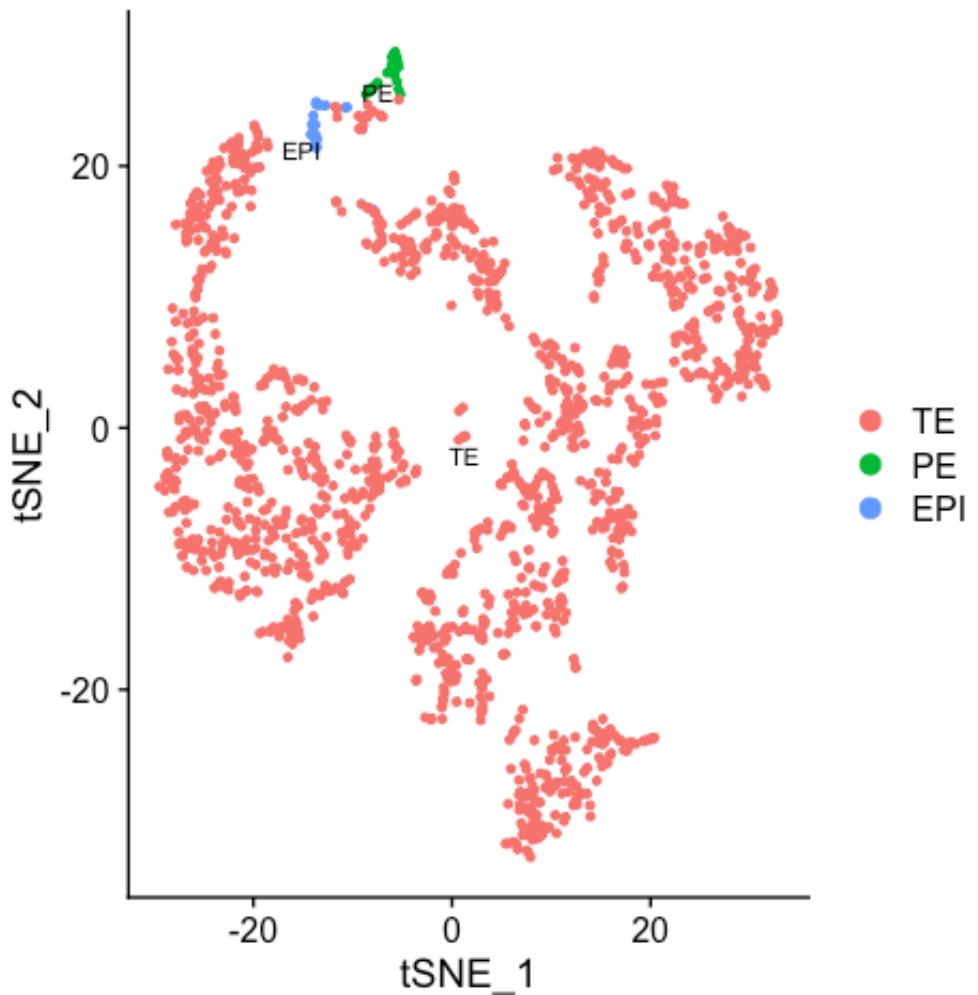
```
cells.ident <- FetchData(all_stage_new, vars="ident")

FeaturePlot(all_stage_new,
            cells=rownames(cells.ident),
            features = c("SLC7A2", "RAB31", "GRHL1", "FYB", "TEAD1", "DLX3",
            "HSD17B1", "HSD3B1", "CLDN4", "ABCG2", "WLS", "MBNL3", "MPP1"), #TE
            label = TRUE)
FeaturePlot(all_stage_new,
            cells=rownames(cells.ident),
            features = c("GPX2", "PDGFRA", "APOA1", "ADD3", "APOE", "GATA4",
            "SPARC", "CKB", "GJA1", "MARCKS", "NID2", "CNN3"),
            label = TRUE) #PE
FeaturePlot(all_stage_new,
            cells=rownames(cells.ident),
            features = c("POU5F1", "NANOG", "SOX2", "DPPA5", "IFITM1", "MEG3",
            "KHDC3L", "TDGF1", "DPPA4", "IFITM3"),
            label = TRUE) #EPI
```





The labeling results are then shown in the cluster results based on the three different cell types:

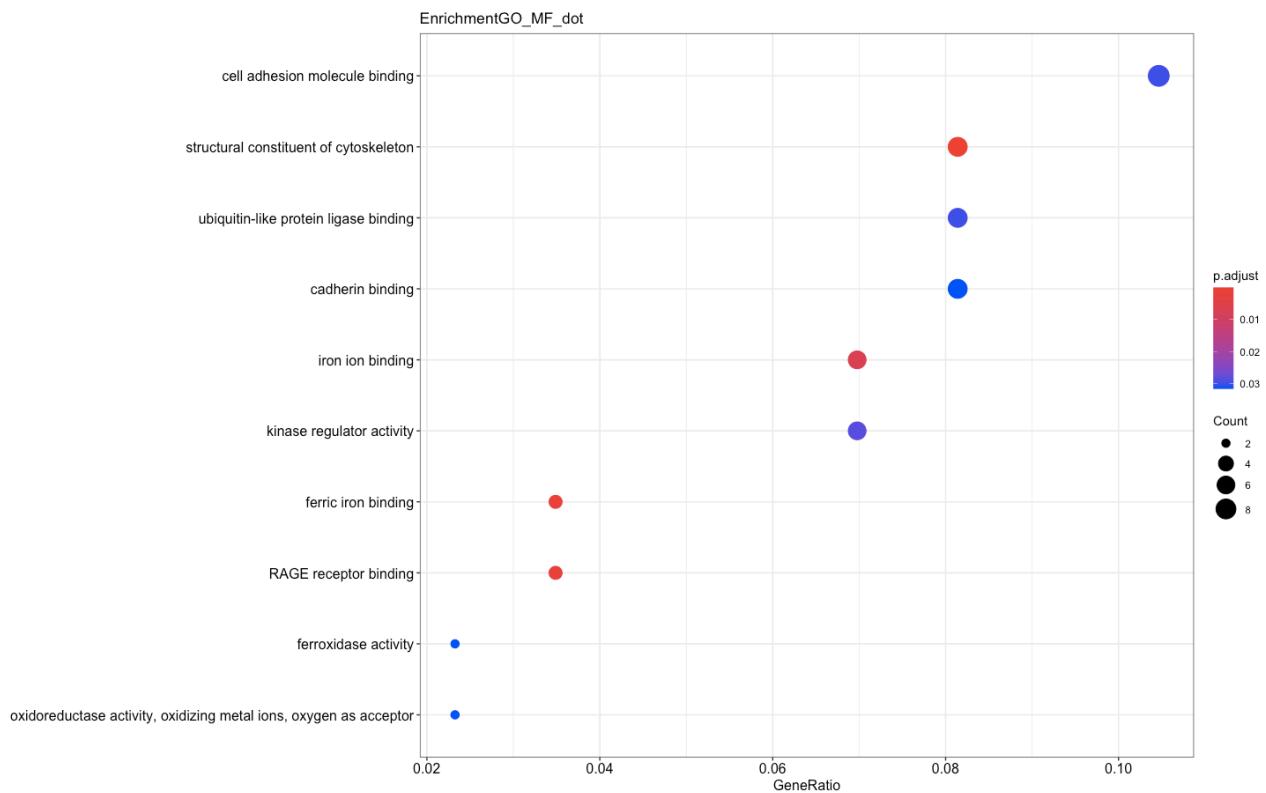


downstream GO analysis

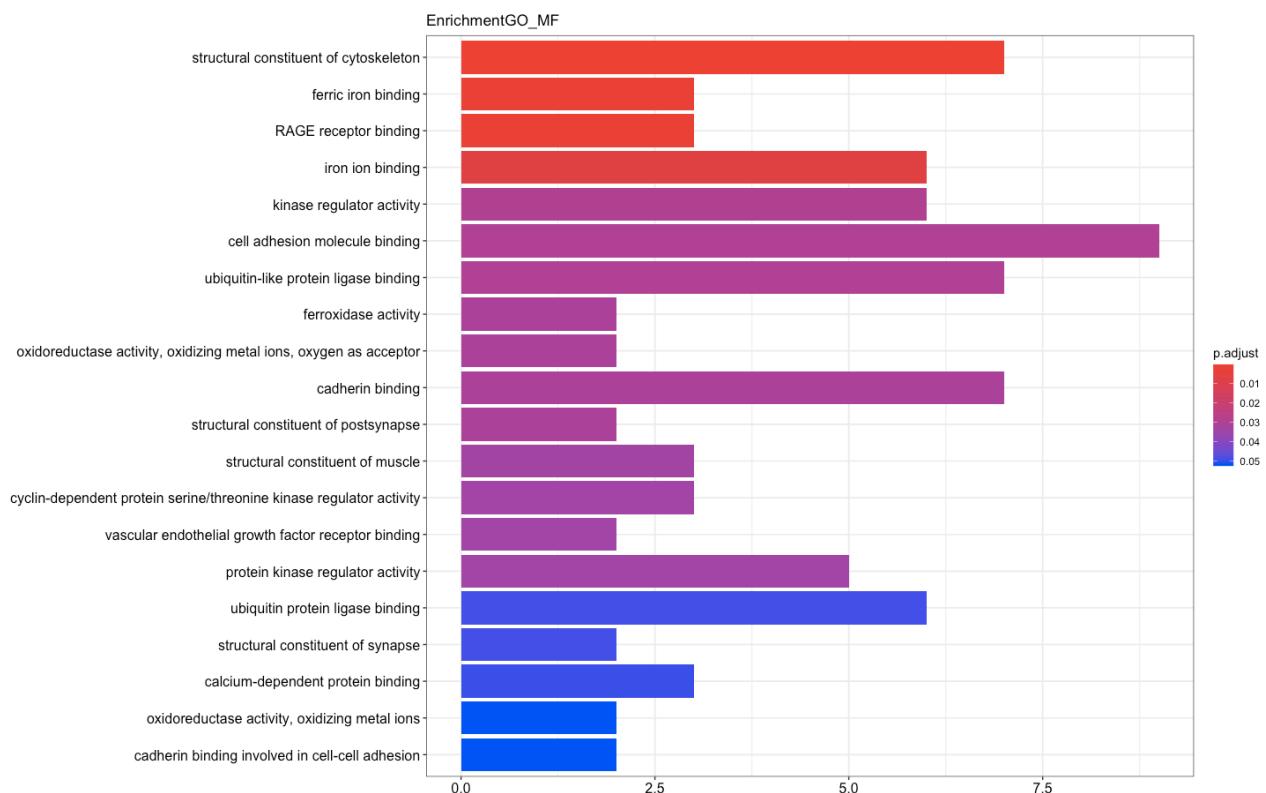
```

BiocManager::install("DOSE")
BiocManager::install("topGO")
BiocManager::install("clusterProfiler")
BiocManager::install("pathview")
library(DOSE)
library(org.Hs.eg.db)
library(topGO)
library(clusterProfiler)
library(pathview)
x <- all_stage_new_top100
test = bitr(x, #your gene list
            fromType="SYMBOL", #SYMBOL format
            toType="ENTREZID", # ENTERZID format
            OrgDb="org.Hs.eg.db") #human database
head(test,2)
ggo <- groupGO(gene = test$ENTREZID, OrgDb = org.Hs.eg.db, ont = "CC",level =
3,readable = TRUE)
genelist <- all_stage_new_genes
genelist = bitr(genelist, #gene list
                fromType="SYMBOL", #SYMBOL format
                toType="ENTREZID", #ENTERZID format
                OrgDb="org.Hs.eg.db") #human database
ego_ALL <- enrichGO(gene = test$ENTREZID,
                      #universe = genelist, #Background gene set
                      OrgDb = org.Hs.eg.db,
                      #keytype = 'ENSEMBL',
                      ont = "ALL", #choose fromm ALL CC BP MF
                      pAdjustMethod = "BH", #chose one method from 'holm",
"hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
                      pvalueCutoff = 1, #The p-value will filter out a lot, so
set pvalueCutoff = 1 to output all the results
                      qvalueCutoff = 1,
                      readable = TRUE) #turn Gene ID to gene Symbol
head(ego_ALL)
ego_MF <- enrichGO(gene = test$ENTREZID, OrgDb = org.Hs.eg.db,ont = "MF",
pAdjustMethod = "BH",pvalueCutoff = 1,qvalueCutoff = 1,readable = FALSE)
head(ego_MF)
dotplot(ego_MF,title="EnrichmentGO_MF_dot")

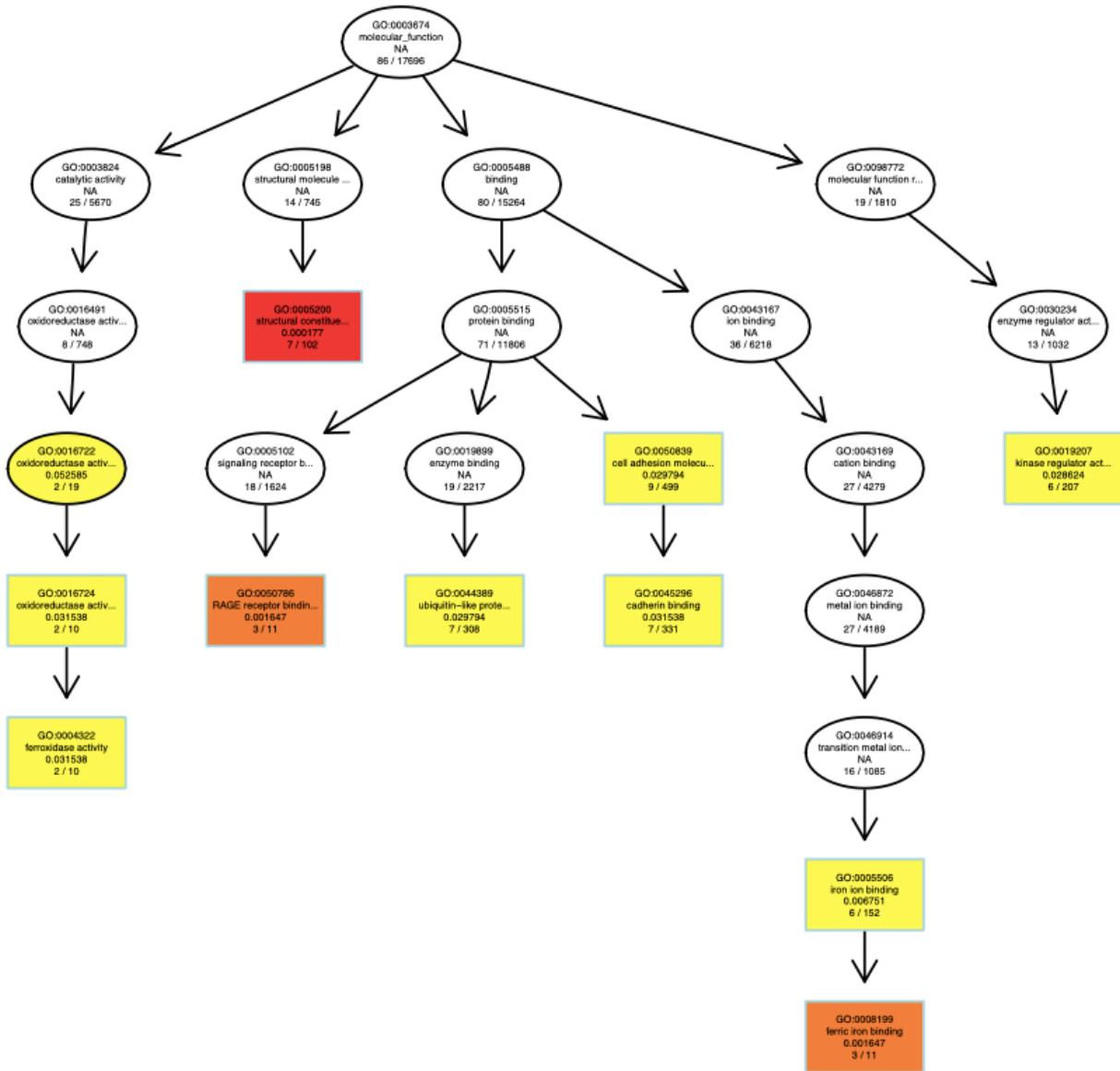
```



```
barplot(ego_MF, showCategory=20, title="EnrichmentGO_MF")
```



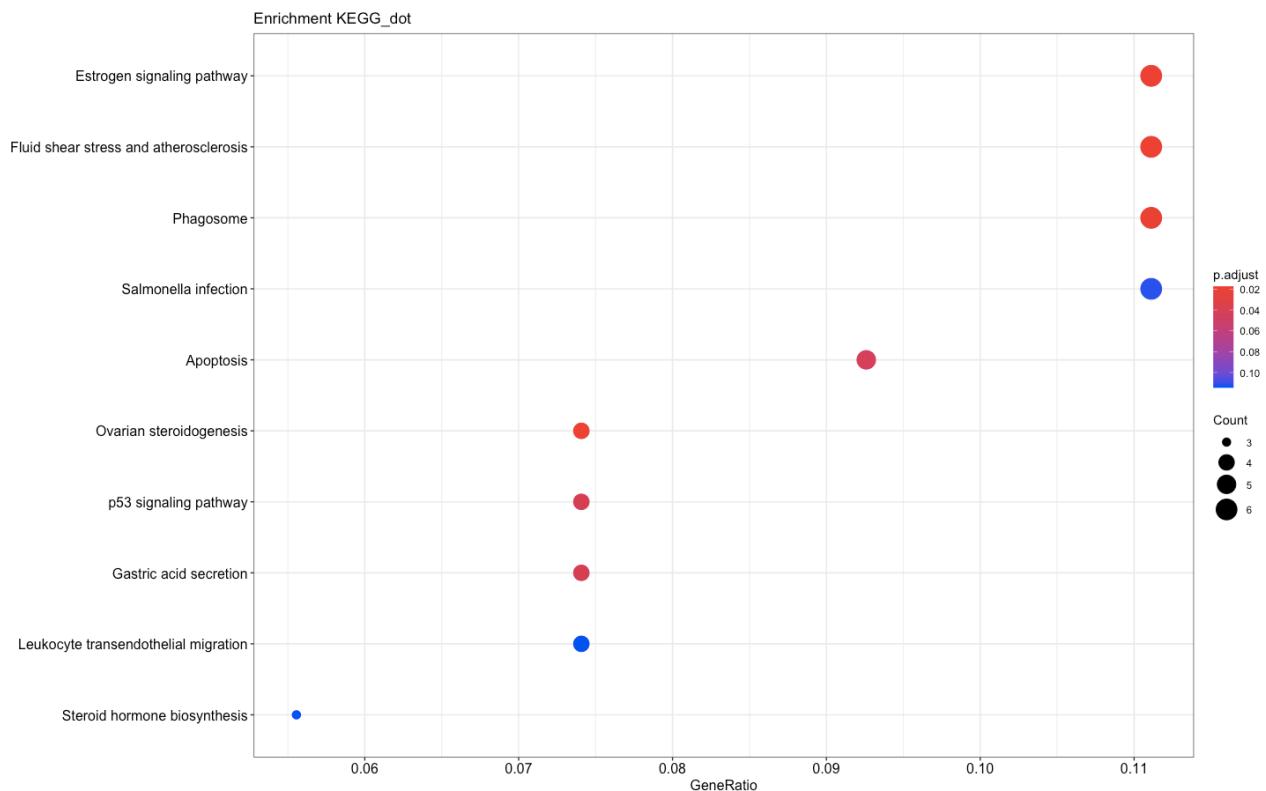
```
plotGOgraph(ego_MF)
```



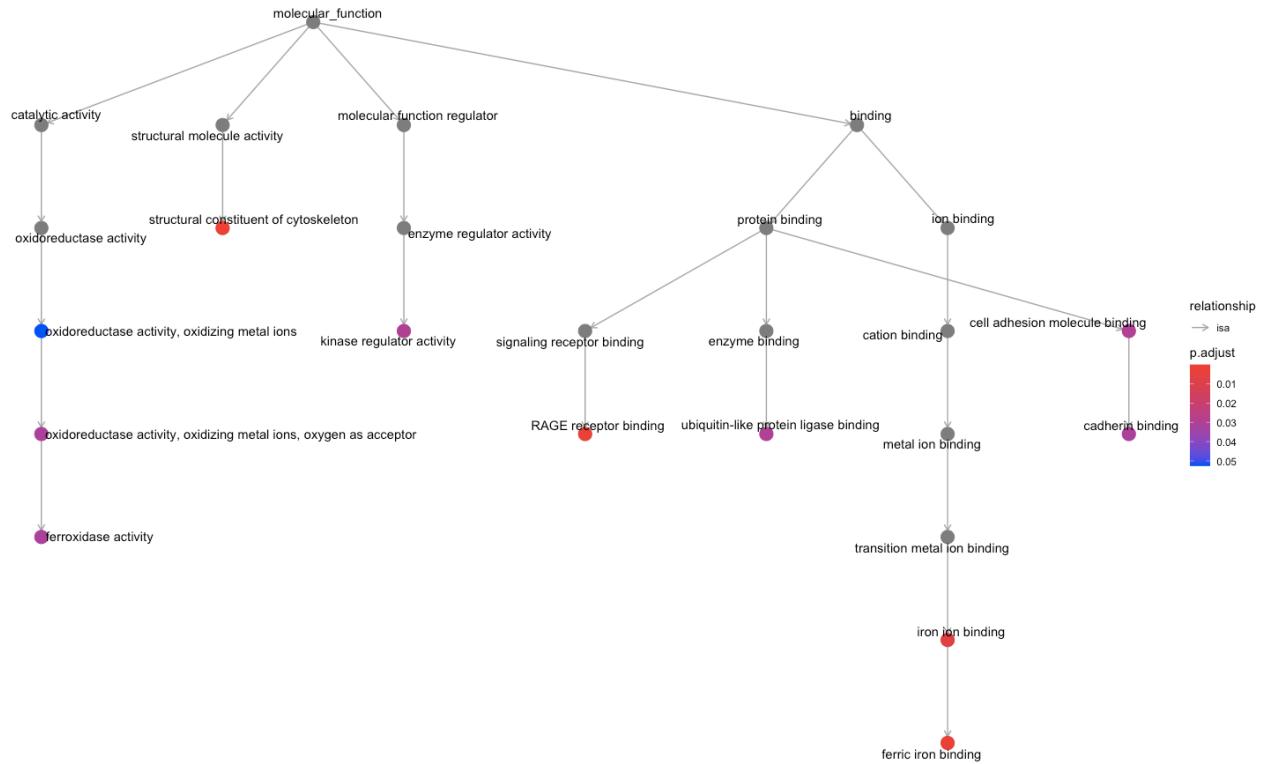
```

ego_MF.fil <- simplify(ego_MF, cutoff=0.05, by="pvalue", select_fun=min)
kk <- enrichKEGG(gene = test$ENTREZID,
                  organism = 'hsa', #KEGG can use organism = 'hsa'
                  pvalueCutoff = 1)
dotplot(kk,title="Enrichment KEGG_dot")

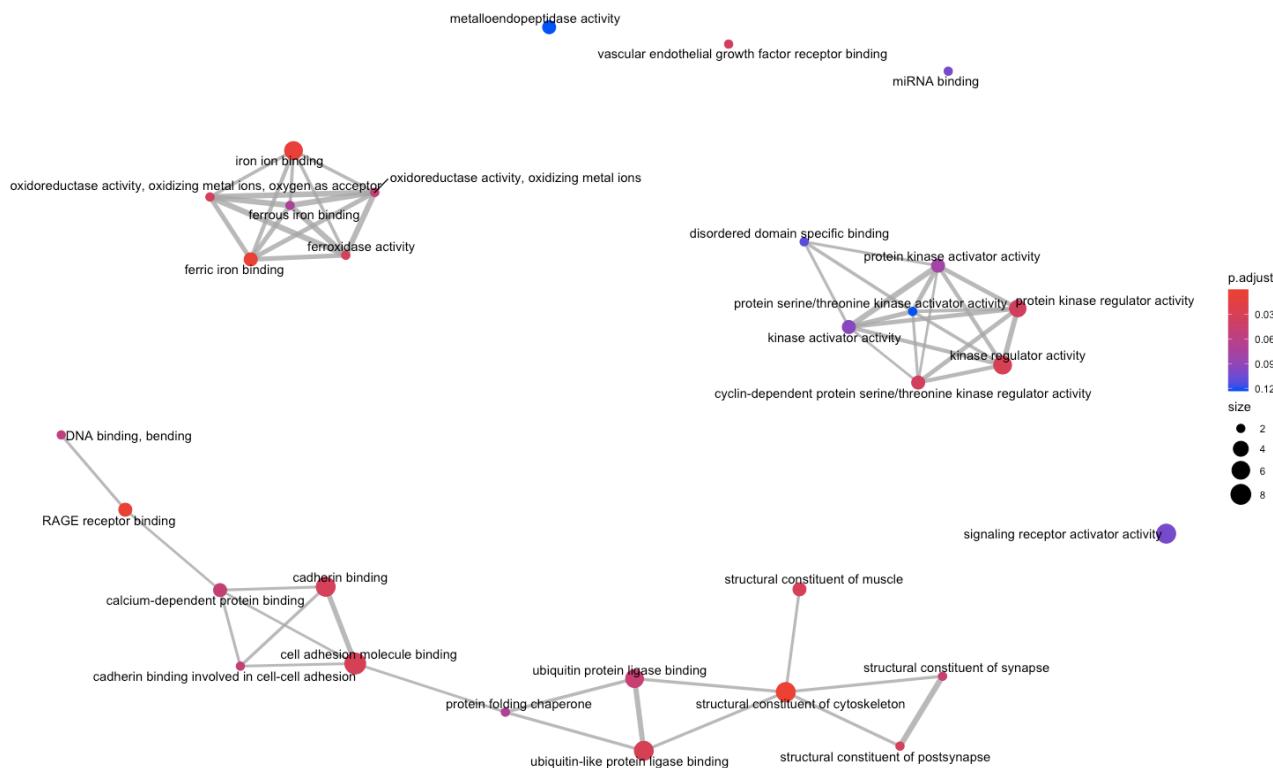
```



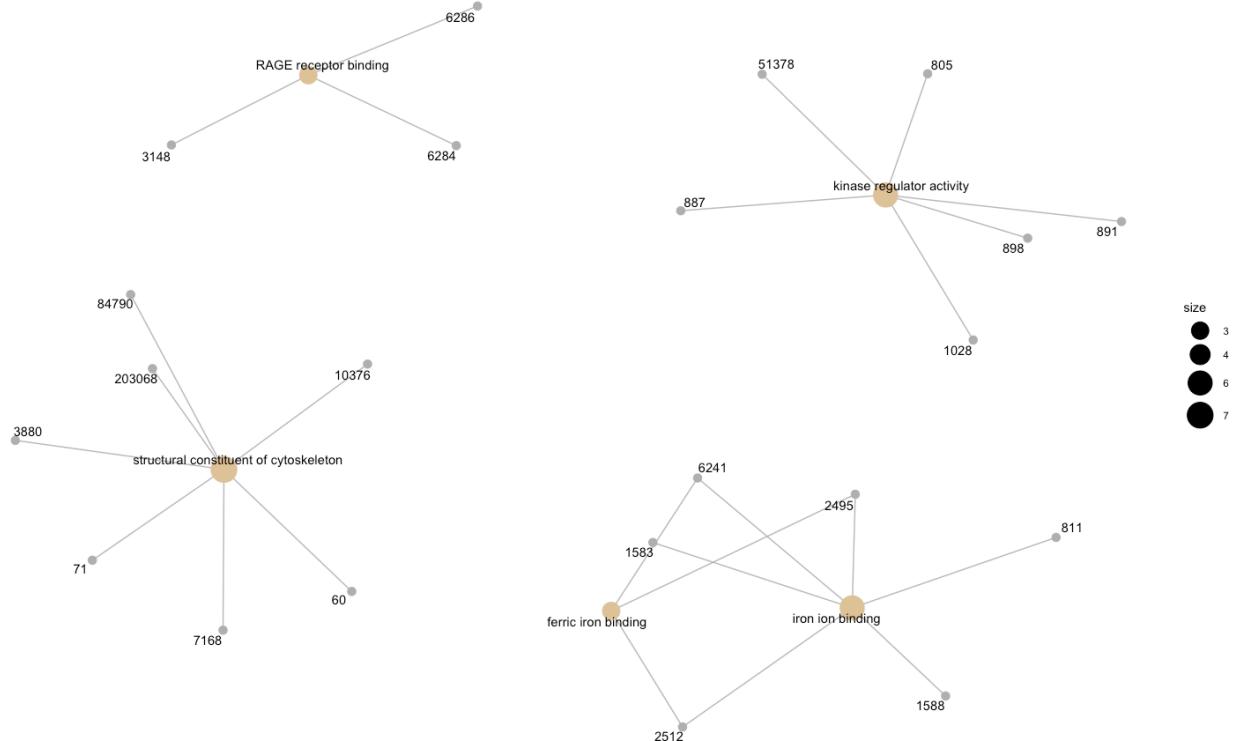
```
write.csv(summary(kk), "/Users/keqinliu/studying_document/bioinfo/genes_involved_in_embryo_irreversible_development_analysis_project/human_embryoImplantation_rawdata/all_stage_new/KEGG-enrich.csv", row.names = FALSE)
hsa04915 <- pathview(gene.data = all_stage_new_genes,
                      pathway.id = "hsa04915",
                      species = "hsa",
                      limit = list(gene=max(abs(test)), cpd=1))
goplot(ego_MF)
```



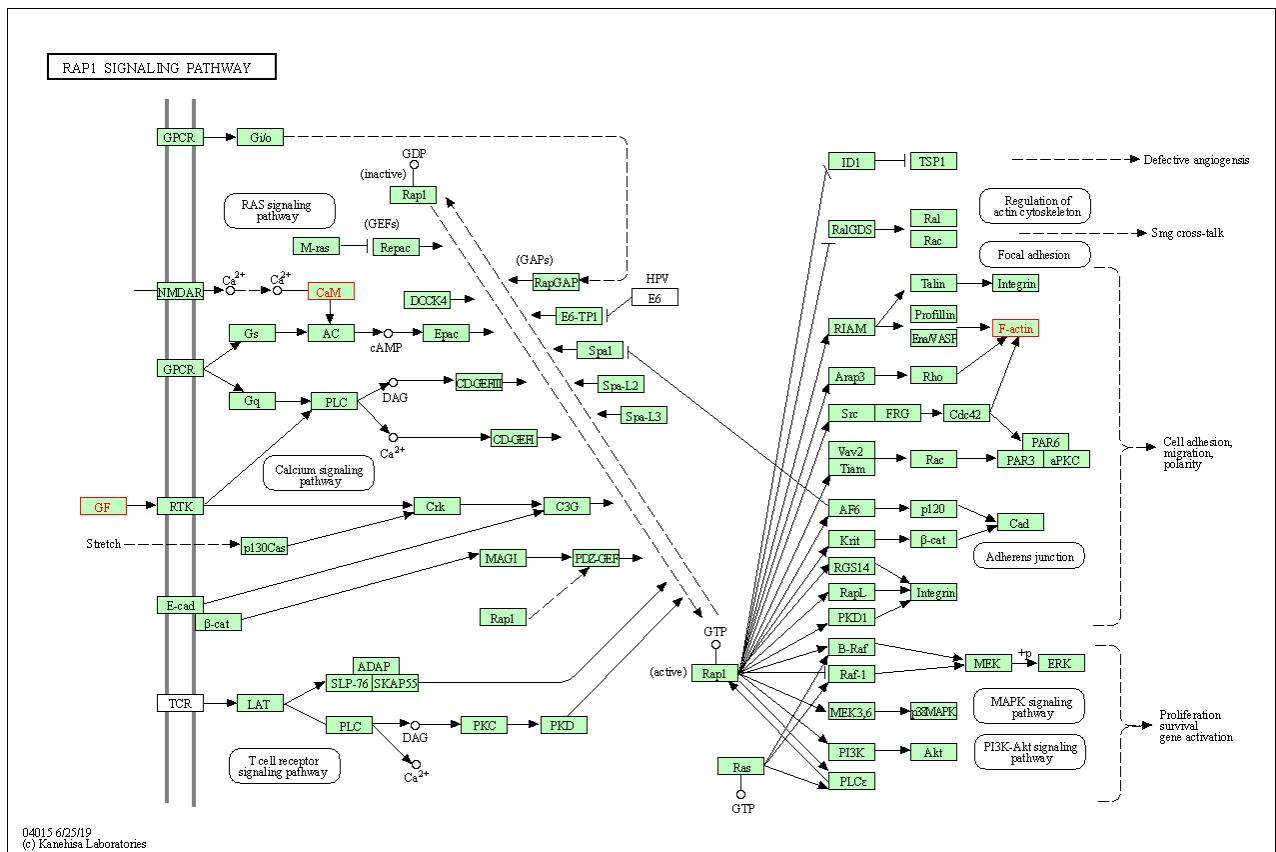
```
emapplot(ego_MF, showCategory = 30)
```

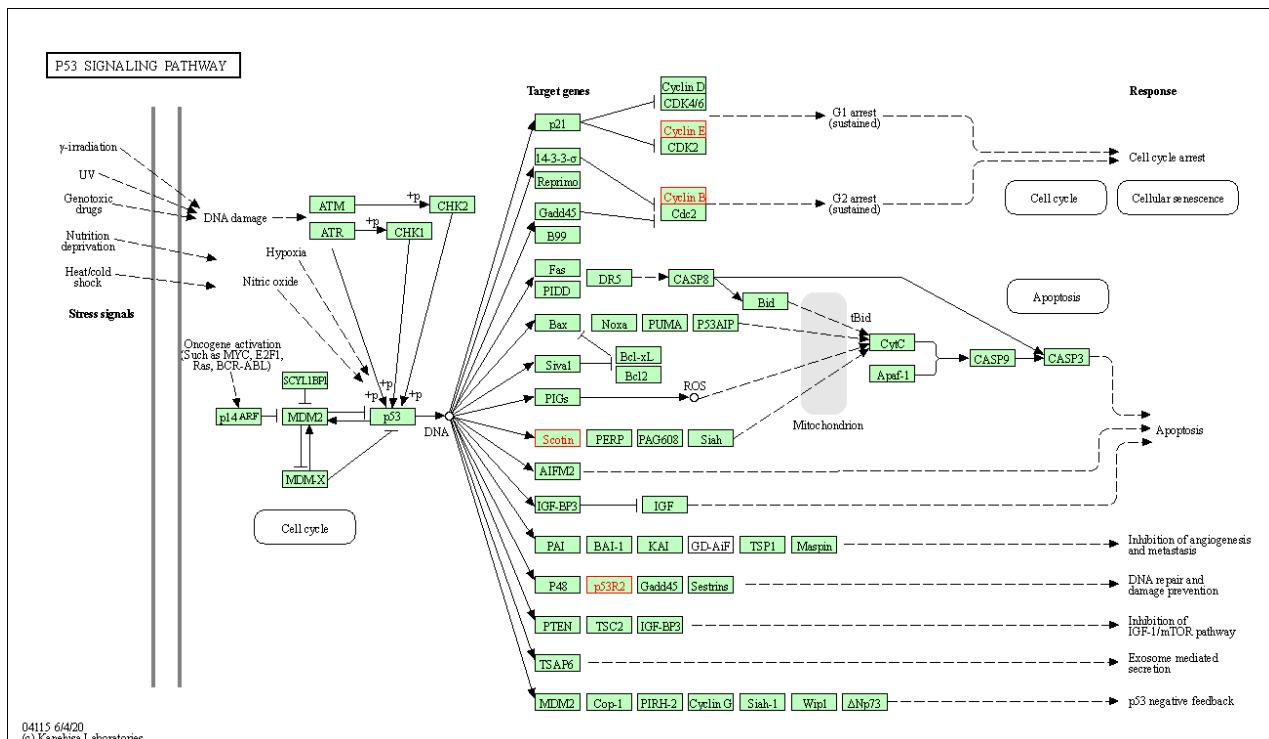


```
cnetplot(ego_MF, showCategory = 5)
```

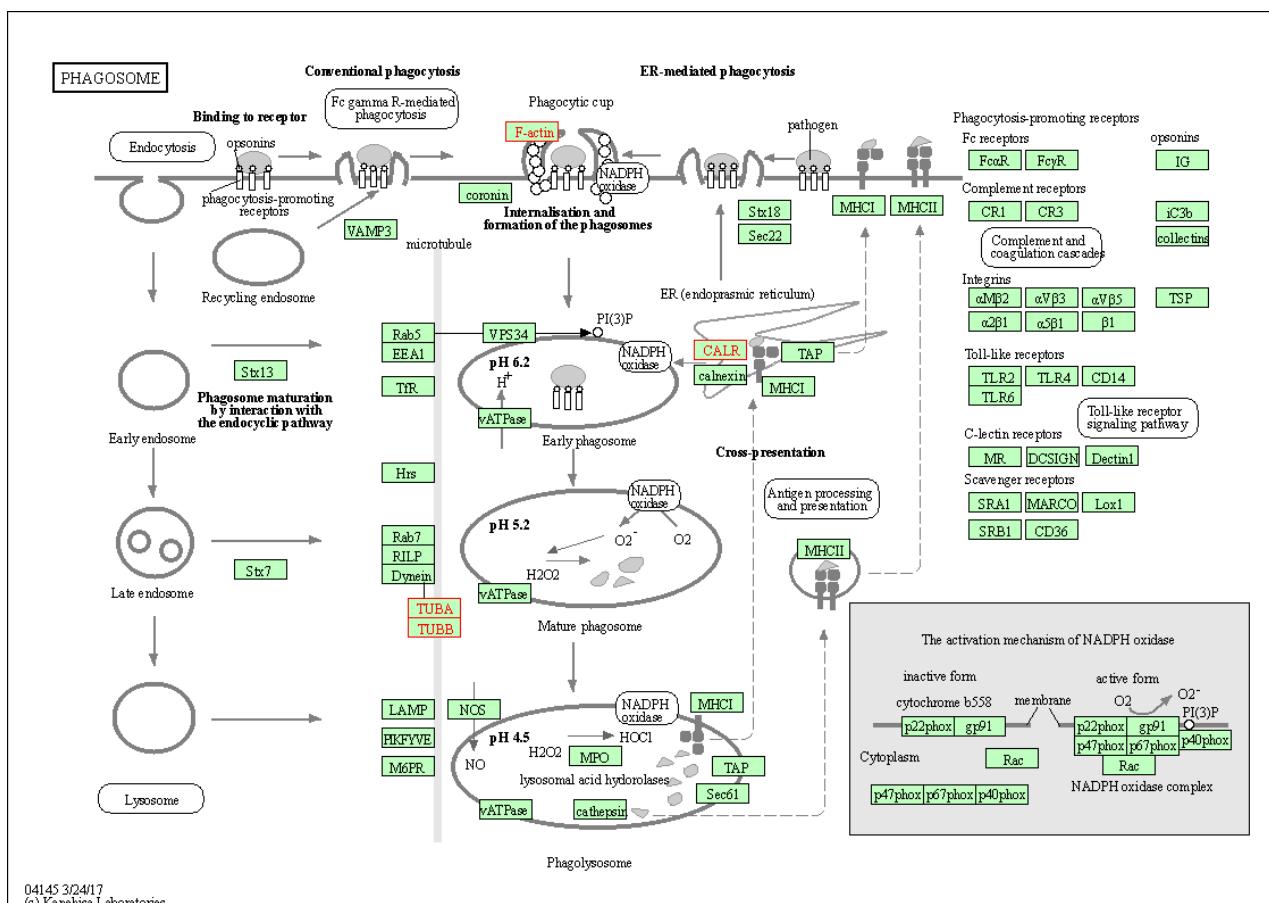


```
oragnx <- setReadable(ego_ALL, 'org.Hs.eg.db', 'ENTREZID')
browseKEGG(kk, 'hsa04015') # hsa xxxxx
```





0415 64420
(c) Kanehisa Laboratories



0415 3/24/17
(c) Kanehisa Laboratories

