绘制超级漂亮的单细胞marker基因气泡图

今天小编为大家介绍一款绘制单细胞marker基因气泡图的工具scDotPlot。在绘制单细胞marker基因气泡图时相信大家主要用的就是两个函数。如果是使用Seurat的同学，估计用的是DotPlot这个Seurat包自带的函数，如果使用的是SingleCellExperiment框架分析单细胞分析的同学估计使用的是scater包提供的plotDots函数。

但是这两个函数有几个不足之处:

1. 不能对行或者列进行聚类。
2. 不能对行和列进行注释。
3. 如果想使用基因表达的z-score值绘制气泡图需要提前对数据进行处理。
4. 不能设置一个阈值，使低于这个阈值的基因表达水平设置为零。

而scDotPlot函数就可以完美解决上述问题。下面就让我们一起来学习如何使用吧！

# 加载需要的R包

library(magrittr) # 使用该包提供的管道符  
library(scDotPlot)  
library(ggsci)  
library(scRNAseq)  
library(scuttle)  
library(scran)  
library(batchelor)  
library(purrr)

library(AnnotationDbi)

# 加载需要的数据

今天我们使用SingleCellExperiment框架为大家展示如何使用scDotPlot函数绘制单细胞基因表达气泡图。

sce <- ZeiselBrainData()  
  
# 查看样本信息  
sce %>% colData %>% head

DataFrame with 6 rows and 9 columns  
 tissue group # total mRNA mol well sex  
 <character> <numeric> <numeric> <numeric> <numeric>  
1772071015\_C02 sscortex 1 21580 11 1  
1772071017\_G12 sscortex 1 21748 95 -1  
1772071017\_A05 sscortex 1 31642 33 -1  
1772071014\_B06 sscortex 1 32916 42 1  
1772067065\_H06 sscortex 1 21531 48 1  
1772071017\_E02 sscortex 1 24799 13 -1  
 age diameter level1class level2class  
 <numeric> <numeric> <character> <character>  
1772071015\_C02 21 0.00 interneurons Int10  
1772071017\_G12 20 9.56 interneurons Int10  
1772071017\_A05 20 11.10 interneurons Int6  
1772071014\_B06 21 11.70 interneurons Int10  
1772067065\_H06 25 11.00 interneurons Int9  
1772071017\_E02 20 11.90 interneurons Int9

# 筛选掉没需注释信息的细胞  
sce <- sce[, sce$level2class != "(none)"]

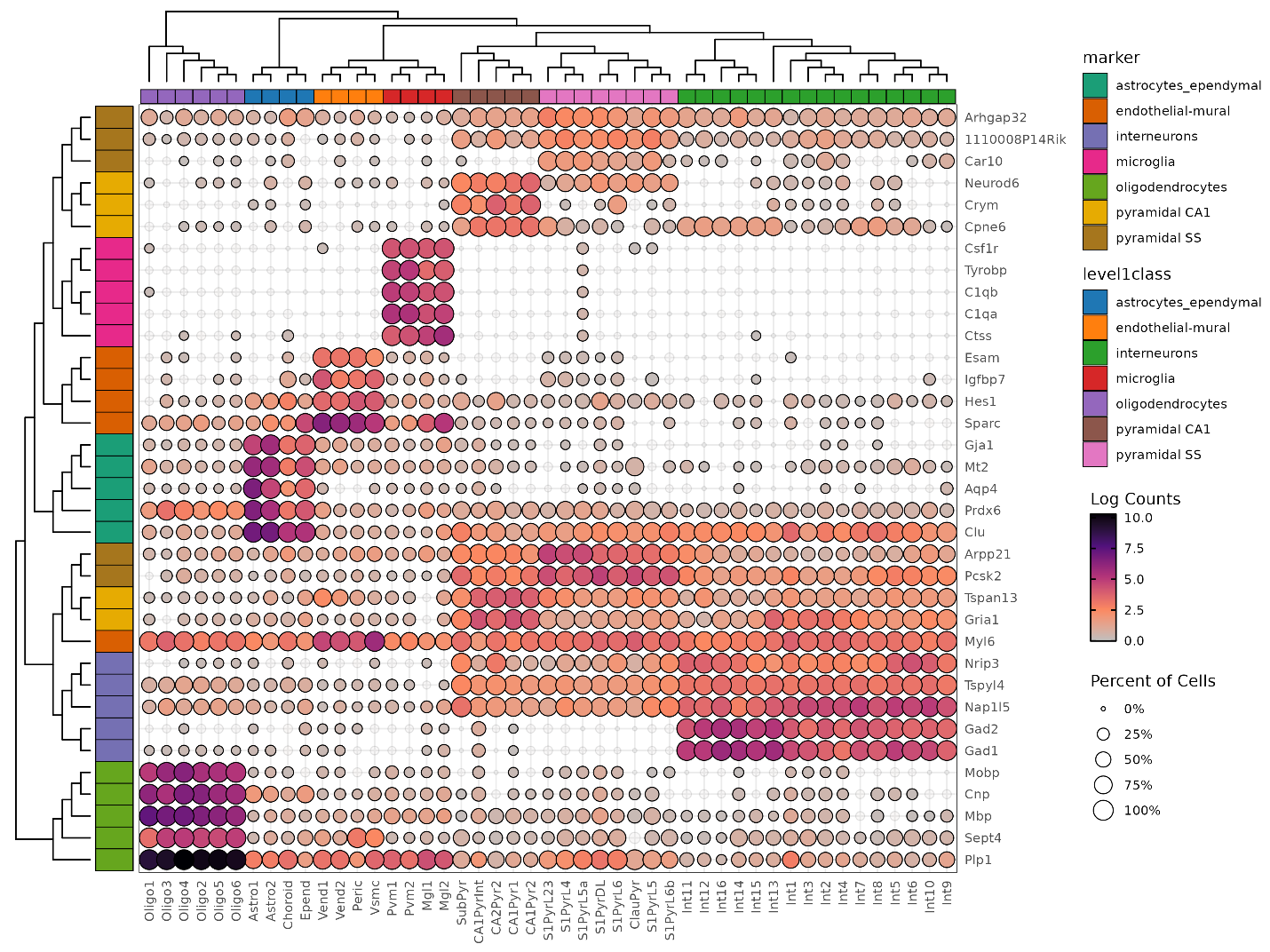
# 向SingleCellExperiment中添加marker信息

# 对sce进行处理  
set.seed(101)  
qc <- quickCluster(sce)  
sce <- computePooledFactors(sce, clusters = qc)  
sce <- logNormCounts(sce, size.factors = sizeFactors(sce))  
  
# 鉴定每种细胞的marker基因  
markers <- findMarkers(x = sce,   
 group = sce$level1class,   
 test.type = "wilcox",   
 pval.type = "all",   
 min.prop = 0.25,   
 direction = "up",   
 lfc = 0.5)  
  
# 把marker信息添加到rowData数据框里面   
# 这里面添加的信息就是每行的基因是属于哪个细胞的marker信息   
cell\_markers <- lapply(markers, function(x){  
 x %>% rownames %>% head(5)  
}) %>% unlist2  
  
rowData(sce)$marker <- cell\_markers[match(rownames(sce), cell\_markers)] %>% names

# 查看添加的marker信息  
sce %>% rowData %>% .$marker %>% table

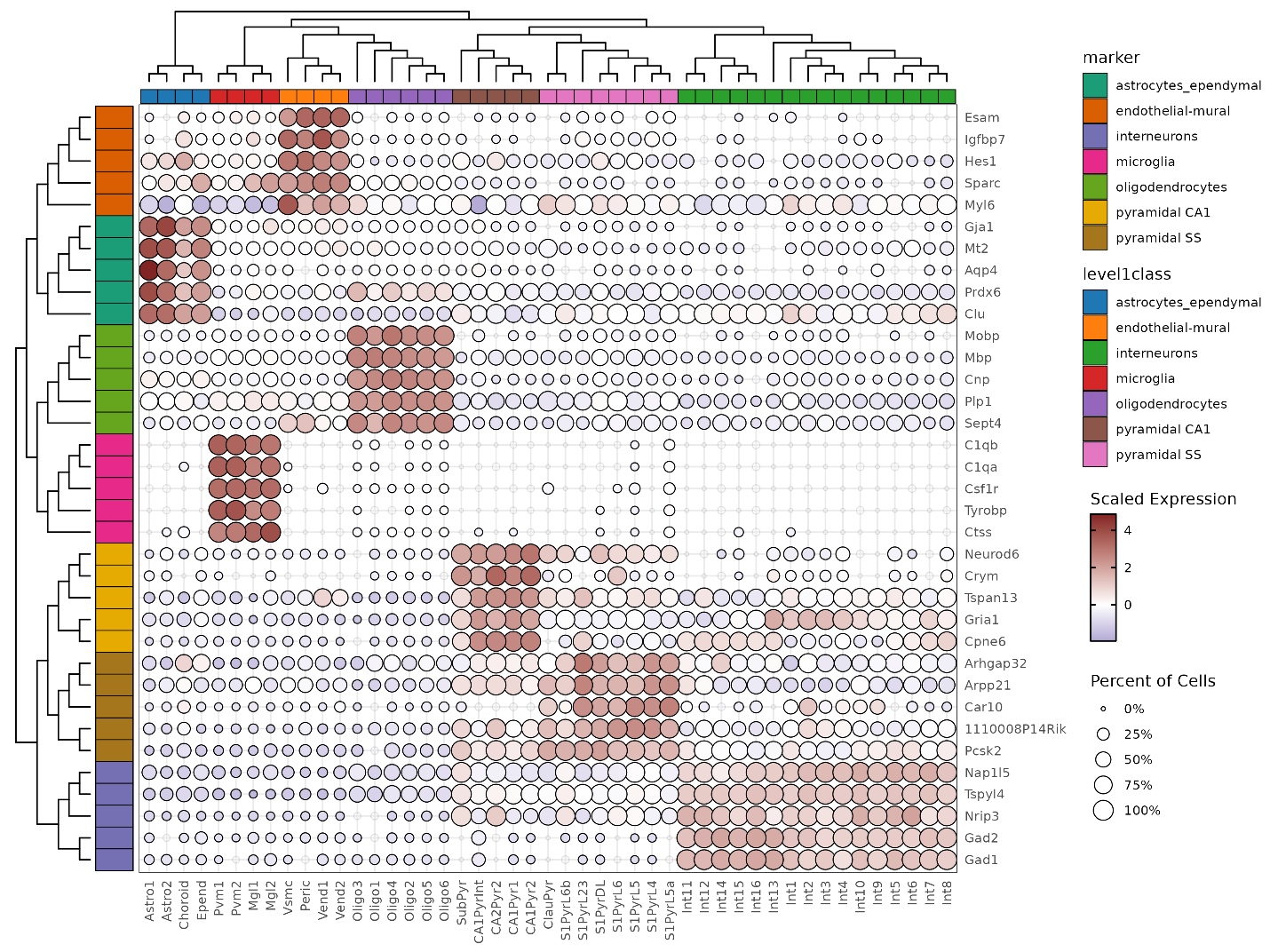
# 使用标准化后的logcounts数据绘制marker基因气泡图

scDotPlot(object = sce,  
 features = unique(cell\_markers), # marker基因  
 group = "level2class", # 细胞统计的单位  
 groupAnno = "level1class", # 列注释信息，colData(sce)中的某一列  
 featureAnno = "marker", # 行注释信息，rowData(sce)中的某一列  
 clusterRows = TRUE, # 对行进行聚类  
 clusterCOlumns = TRUE, # 对列进行聚类，   
 AverageThreshold = 0.1, # 在某种细胞中低于这个阈值认为这个基因在这种细胞中不表达  
 NumDetectedThreshold = 0.1, # 在某种细胞中低于这个检查率认为在这个细胞中不表达  
 groupLegends = TRUE, # 是否显示列注释的图例  
 featureLegends = TRUE #是否显示行注释的图例  
 )



# 使用z-score绘制气泡图

scDotPlot(object = sce,   
 features = unique(cell\_markers),   
 group = "level2class",   
 groupAnno = "level1class",   
 featureAnno = "marker",   
 clusterRows = TRUE,   
 clusterColumns = TRUE,   
 # AverageThreshold = 0.01,   
 NumDetectedThreshold = 0.05,  
 groupLegends = TRUE,   
 featureLegends = TRUE,   
 scale = TRUE # 使用z-score绘制气泡图  
)



好了今天小编就为大家介绍到这里啦，有什么问题大家可以在下面留言交流。本公众号后续还会更新更多空间转录组、空间蛋白质组、空间代谢组以及单细胞方面的内容。本公众号由上海萤锐科技有限公司运营，上海萤锐科技主要专注与单细胞测序、空间转录组、空间代谢组、空间蛋白质组以及宏基因组方向，欢迎大家交流咨询。本公众好后续会不定期更新更多相关内容，欢迎大家关注公众号留言讨论。

