sravg_signac

Qingnan Liang

5/17/2022

Introduction

Here is a demonstration of how to use the package SRAVG to create a Seurat object with cells grouped and averaged.

Load libraries

```
library(Signac)
library(Seurat)
library(EnsDb.Hsapiens.v86)
library(BSgenome.Hsapiens.UCSC.hg38)
library(SRAVG)
```

Use the pbmc3k-multiome dataset from the 10x website. This data is preprocessed using default Signac parameters

```
data <- readRDS("C:/Users/liang/work/34_gene_correlation_simpson/LinkPeak/pbmc3k.rds")
print(data@assays$RNA)
```

```
## Assay data with 36601 features for 2398 cells
## First 10 features:
##  MIR1302-2HG, FAM138A, OR4F5, AL627309.1, AL627309.3, AL627309.2,
## AL627309.5, AL627309.4, AP006222.2, AL732372.1
```

```
print(data@assays$peaks)
```

```
## ChromatinAssay data with 91450 features for 2398 cells
## Variable features: 91386
## Genome:
## Annotation present: TRUE
## Motifs present: FALSE
## Fragment files: 1
```

```
print(data@reductions)
```

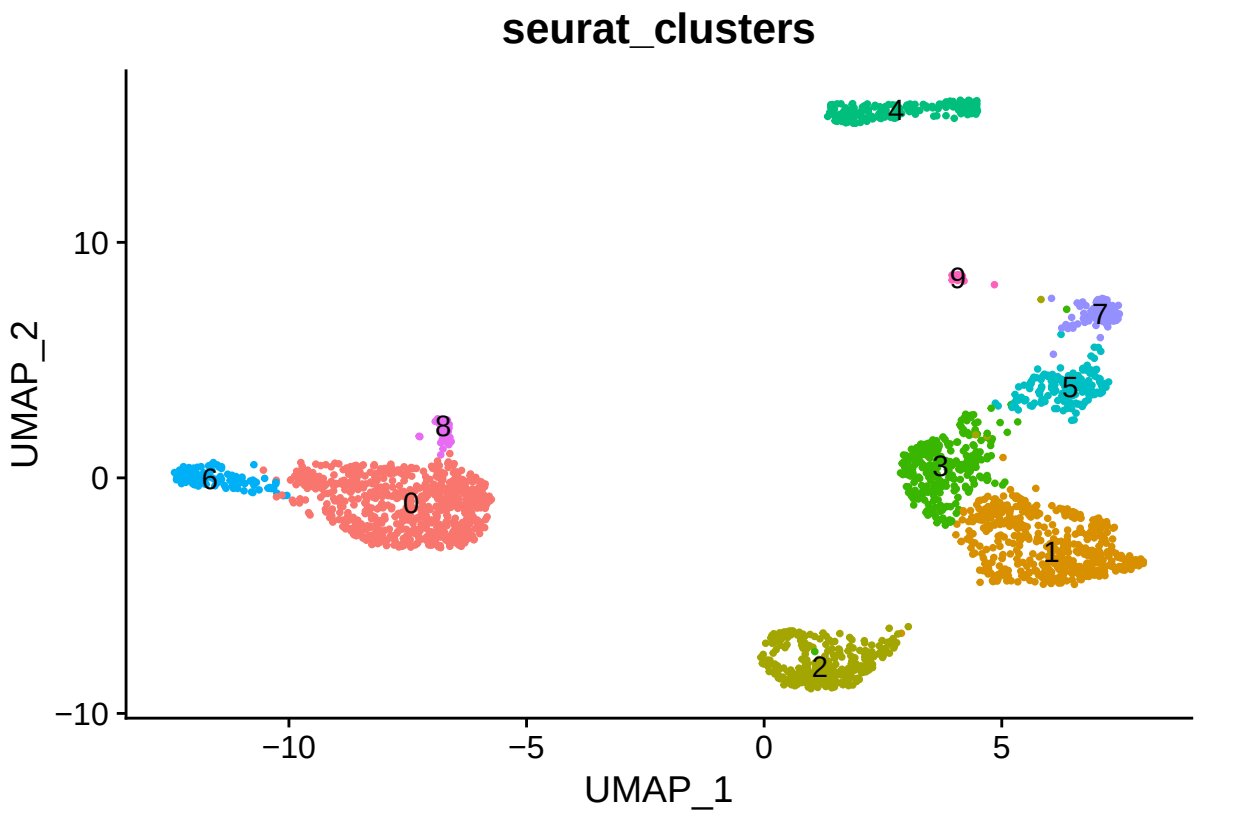
```
## $pca
## A dimensional reduction object with key PC_
```

```
##  Number of dimensions: 50
##  Projected dimensional reduction calculated:  FALSE
##  Jackstraw run: FALSE
##  Computed using assay: SCT
##
## $lsi
## A dimensional reduction object with key LSI_
##  Number of dimensions: 50
##  Projected dimensional reduction calculated:  FALSE
##  Jackstraw run: FALSE
##  Computed using assay: peaks
##
## $umap
## A dimensional reduction object with key UMAP_
##  Number of dimensions: 2
##  Projected dimensional reduction calculated:  FALSE
##  Jackstraw run: FALSE
##  Computed using assay: RNA
```

"RNA" and "peaks" are the two assays we want to average; "pca" will be used as the low-dimensional coordinates for clustering cells to meta-cells.

Visualize the clustering at single-cell level:

```
DimPlot(data, group.by = "seurat_clusters", label = TRUE) + NoLegend()
```



Run SRAVG

Here we use 'pca' as the dimension reduction to evaluate the distances between cells, and top 10 PCs are used (dr_dims). We try to form meta-cells by averaging 10 cells to 1. The group will be within each individual 'seurat_clusters'. We also average two other columns in the meta.data and transfer to the output seurat, which are 'nCount_RNA', and 'nFeature_RNA'. At this time we only support numerical columns for 'extra_meta'.

```
start_time <- Sys.time()

data_avg <- sravg(object = data, dr_key = "pca", dr_dims = 10,
    group_size = 10, group_within = "seurat_clusters", peak_assay = "peaks",
    peak_slot = "data", extra_meta = c("nCount_RNA", "nFeature_RNA",
        "nCount_ATAC", "nFeature_ATAC"))
end_time <- Sys.time()
end_time - start_time
```

```
## Time difference of 27.6493 secs
```

The output is a Seurat object

```
print(data_avg@assays$RNA)
```

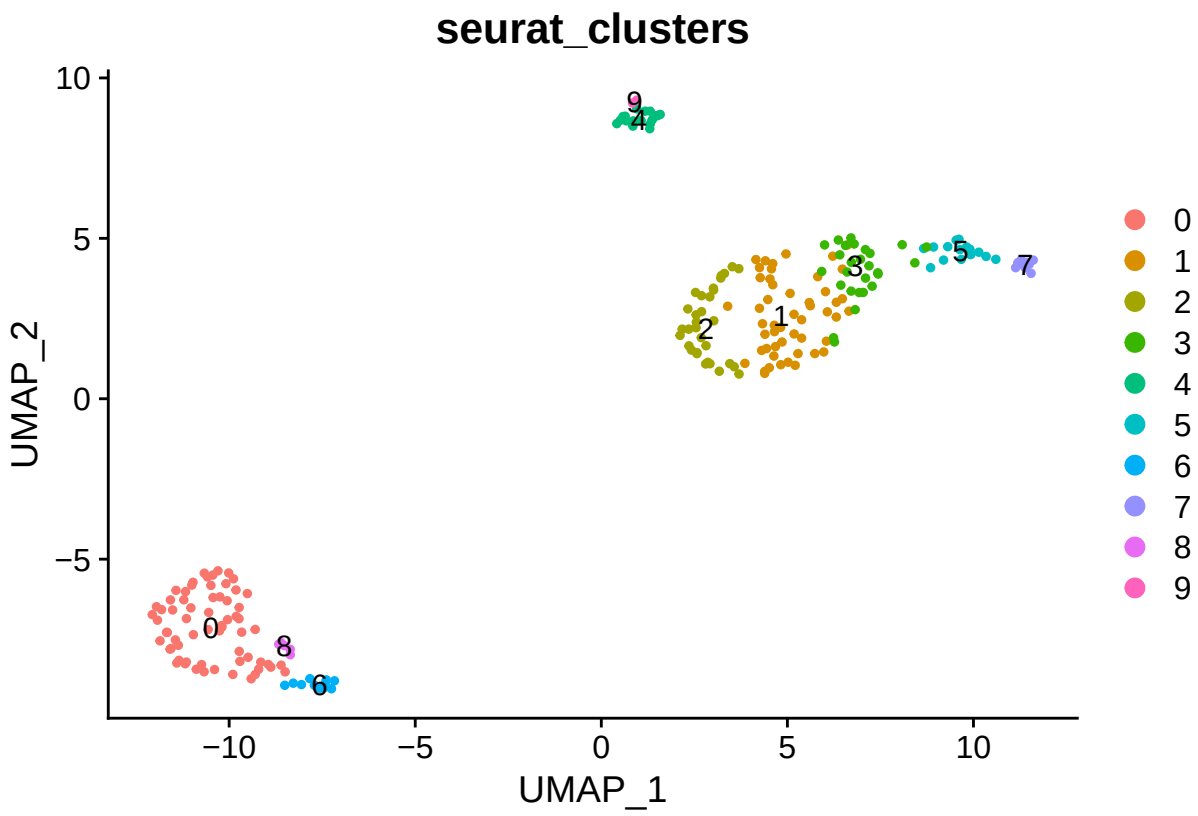
```
## Assay data with 36601 features for 235 cells
## First 10 features:
##  MIR1302-2HG, FAM138A, OR4F5, AL627309.1, AL627309.3, AL627309.2,
## AL627309.5, AL627309.4, AP006222.2, AL732372.1
```

```
print(data_avg@assays$peaks)
```

```
## ChromatinAssay data with 91450 features for 235 cells
## Variable features: 0
## Genome:
## Annotation present: TRUE
## Motifs present: FALSE
## Fragment files: 0
```

We can still run UMAP on the data_avg object. The 'pca' in this averaged object is calculated with averaging the original 'pca' coordinates (for each meta-cell).

```
data_avg <- RunUMAP(data_avg, dims = 1:10, verbose = FALSE)
DimPlot(data_avg, group.by = "seurat_clusters", label = T)
```



Compute the sparsity of matrix (proportion of zeros) before and after averaging

```
sparsity <- function(matrix) {
    sparsity <- sum(matrix == 0)/(dim(matrix)[1] * dim(matrix)[2])
    return(sparsity)
}

print(sparsity(data@assays$RNA@counts))
```

```
## [1] 0.9465753
```

```
print(sparsity(data_avg@assays$RNA@counts))
```

```
## [1] 0.7970439
```

```
print(sparsity(data@assays$peaks@counts))
```

```
## [1] 0.9174988
```

```
print(sparsity(data_avg@assays$peaks@counts))
```

```
## [1] 0.6765056
```