Seurat - Clustering Tutorial

Dennis Dimitri Krutkin

2023-06-15

```
#Load necessary libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(Seurat)
```

```
## Attaching SeuratObject
```

```
library(patchwork)
```

Import 10X dataset

```
# Load the PBMC data set:
pbmc.data <- Read10X(data.dir = "./filtered_gene_bc_matrices/hg19/")

# Initialize the Seurat object with the raw (non-normalized data):
pbmc <- CreateSeuratObject(counts = pbmc.data, project = "pbmc3k", min.cells = 3, min.features = 200)
```

```
## Warning: Feature names cannot have underscores ('_'), replacing with dashes
## ('-')
```

```
pbmc
```

```
## An object of class Seurat
## 13714 features across 2700 samples within 1 assay
## Active assay: RNA (13714 features, 0 variable features)
```

Examine the loaded dataset

```
# Examine a 3 genes in the first 25 cells:
pbmc.data[c("CD3D", "TCL1A", "MS4A1"), 1:25]
```

```
## 3 x 25 sparse Matrix of class "dgCMatrix"
```

```
##    [[ suppressing 25 column names 'AAACATACAACCAC-1', 'AAACATTGAGCTAC-1', 'AAACATTGATCAGC-1' ...
]]
```

```
##
## CD3D  4 . 10 . . 1 2 3 1 . . 2 7 1 . . 1 3 . 2  3 . . . .
## TCL1A . .  . . . . . . 1 . . . . . . . . . . . .  . 1 . . .
## MS4A1 . 6  . . . . . . 1 1 1 . . . . . . . . . 36 1 2 . .
```

The “.” values in the matrix represent 0s (no molecules detected).

Most values in a scRNA-seq matrix are 0 - Seurat uses a sparse-matrix representation whenever possible. This results in significant memory and speed savings for Drop-seq/inDrop/10x data.