

Single-cell RNA-sequencing intro

Koen Van den Berge

Last compiled on 26 November, 2021

Contents

```
suppressPackageStartupMessages(library(scRNAseq))
sce <- MacoskoRetinaData()
```

```
## snapshotDate(): 2021-10-19
```

```
## see ?scRNAseq and browseVignettes('scRNAseq') for documentation
```

```
## loading from cache
```

```
## see ?scRNAseq and browseVignettes('scRNAseq') for documentation
```

```
## loading from cache
```

```
sce
```

```
## class: SingleCellExperiment
## dim: 24658 49300
## metadata(0):
## assays(1): counts
## rownames(24658): KITL TMTC3 ... 1110059M19RIK GM20861
## rowData names(0):
## colnames(49300): r1_GGCCGCAGTCCG r1_CTTGTGCGGGAA ... p1_TAACGCGCTCCT
##   p1_ATTCTTGTTCTT
## colData names(2): cell.id cluster
## reducedDimNames(0):
## mainExpName: NULL
## altExpNames(0):
```

```
class(sce)
```

```
## [1] "SingleCellExperiment"
## attr("package")
## [1] "SingleCellExperiment"
```

```
counts(sce)[1:5,1:5]
```

```
## 5 x 5 sparse Matrix of class "dgMatrix"
##           r1_GGCCGCAGTCCG r1_CTTGTGCGGGAA r1_GCGCAACTGCTC r1_GATTGGGAGGCA
## KITL                      .                .                1                .
## TMTC3                     3                .                .                .
## CEP290                    1                3                .                2
## 4930430F08RIK             2                1                2                .
## 1700017N19RIK             .                .                .                .
##           r1_CCTCCTAGTTGG
## KITL                      .
## TMTC3                     2
## CEP290                    1
## 4930430F08RIK             1
## 1700017N19RIK             .
```

```
head(colData(sce))
```

```
## DataFrame with 6 rows and 2 columns
##           cell.id  cluster
##           <character> <integer>
## r1_GGCCGCAGTCCG r1_GGCCGCAGTCCG      2
## r1_CTTGTGCGGGAA r1_CTTGTGCGGGAA      2
## r1_GCGCAACTGCTC r1_GCGCAACTGCTC      2
## r1_GATTGGGAGGCA r1_GATTGGGAGGCA      2
## r1_CCTCCTAGTTGG r1_CCTCCTAGTTGG     NA
## r1_AGTCAAGCCCTC r1_AGTCAAGCCCTC     NA
```

```
# filter cells
```

```
sce <- sce[,!is.na(colData(sce)$cluster)]
sce
```

```
## class: SingleCellExperiment
## dim: 24658 44808
## metadata(0):
## assays(1): counts
## rownames(24658): KITL TMTC3 ... 1110059M19RIK GM20861
## rowData names(0):
## colnames(44808): r1_GGCCGCAGTCCG r1_CTTGTGCGGGAA ... p1_TAACGCGCTCCT
##   p1_ATTCTTGTCTT
## colData names(2): cell.id cluster
## reducedDimNames(0):
## mainExpName: NULL
## altExpNames(0):
```

- Explore this dataset. What do you think is different to these data as compared to a bulk RNA-seq dataset?
- Try visualizing the structure of this dataset using tools we have worked with before. For example, make a PCA and MDS plot. You can color the cells according to the cluster labels in the `colData`.
- Try visualizing the structure of this dataset using any tool you want.

After trying:

Are you able to recapitulate the structure (e.g., cell type clusters)? What are the issues you are encountering, and **why** do you think these are happening?