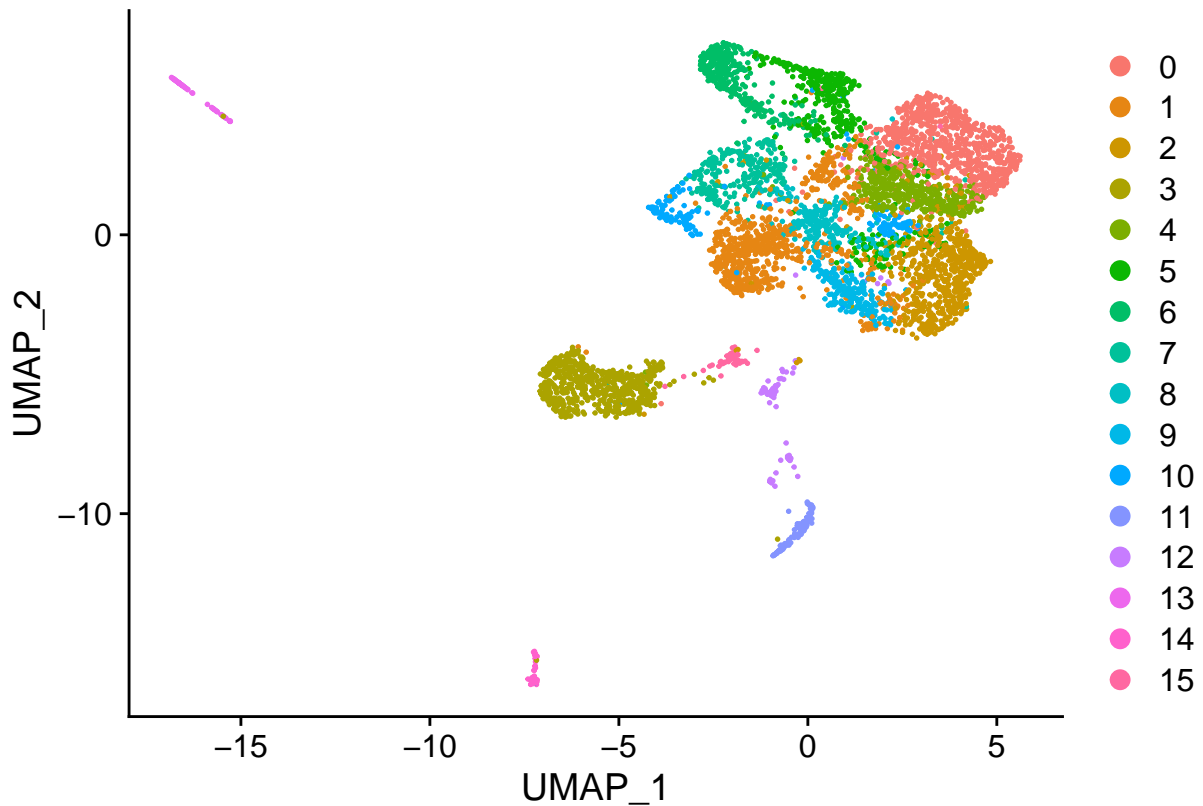# RIG_scRNAseq

```r
# Seurat v3.0.0 and tximport >=1.13.10
library(Seurat)
pbmc <- readRDS(file = "Z:/ResearchHome/ProjectSpace/lucasgrp/HGG_radiation/common/scRNAseq/201110_RIGs_

set.seed(42)
pbmc$replicate <- sample(c("rep1", "rep2"), size = ncol(pbmc), replace = TRUE)
```

```r
# Plot UMAP, coloring cells by cell type (currently stored in object@ident)
DimPlot(pbmc, reduction = "umap")
```



```r
# How many cells are in each cluster
table(Idents(pbmc))
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
## 788 719 676 605 477 410 321 293 231 229 196 112  90  78  60  53
```

```r
# How many cells are in each replicate?
table(pbmc$replicate)
```

```
## 
## rep1 rep2 
## 2666 2672
```

```r
# What proportion of cells are in each cluster?
prop.table(table(Idents(pbmc)))
```

```
## 
##           0           1           2           3           4           5
## 0.147620832 0.134694642 0.126639191 0.113338329 0.089359311 0.076807793
##           6           7           8           9          10          11
## 0.060134882 0.054889472 0.043274635 0.042899963 0.036717872 0.020981641
##          12          13          14          15
## 0.016860247 0.014612214 0.011240165 0.009928812
```

```r
# How does cluster membership vary by replicate?
table(Idents(pbmc), pbmc$replicate)
```

```
## 
##       rep1 rep2
##   0    404  384
##   1    351  368
##   2    342  334
##   3    297  308
##   4    238  239
##   5    207  203
##   6    167  154
##   7    151  142
##   8    106  125
##   9    109  120
##   10    97   99
##   11    54   58
##   12    42   48
##   13    40   38
##   14    29   31
##   15    32   21
```

```r
prop.table(table(Idents(pbmc), pbmc$replicate), margin = 2)
```

```
## 
##             rep1        rep2
##   0  0.151537884 0.143712575
##   1  0.131657914 0.137724551
##   2  0.128282071 0.125000000
##   3  0.111402851 0.115269461
##   4  0.089272318 0.089446108
##   5  0.077644411 0.075973054
##   6  0.062640660 0.057634731
```

```
##    7  0.056639160 0.053143713
##    8  0.039759940 0.046781437
##    9  0.040885221 0.044910180
##    10 0.036384096 0.037050898
##    11 0.020255064 0.021706587
##    12 0.015753938 0.017964072
##    13 0.015003751 0.014221557
##    14 0.010877719 0.011601796
##    15 0.012003001 0.007859281
```

```r
# How can I extract expression matrix for all NK cells (perhaps, to load into another package)
zero.raw.data <- as.matrix(GetAssayData(pbmc, slot = "counts")[, WhichCells(pbmc, ident = "0")])

# Can I create a Seurat object based on expression of a feature or value (MS4A1) in object metadata?
subset(pbmc, subset = MS4A1 > 1)
```

```
## An object of class Seurat
## 43718 features across 7 samples within 2 assays
## Active assay: SCT (21043 features, 3000 variable features)
##  1 other assay present: RNA
##  3 dimensional reductions calculated: pca, harmony, umap
```

```r
# Can I create a Seurat object based on replicate in object metadata?
subset(pbmc, subset = replicate == "rep2")
```

```
## An object of class Seurat
## 43718 features across 2672 samples within 2 assays
## Active assay: SCT (21043 features, 3000 variable features)
##  1 other assay present: RNA
##  3 dimensional reductions calculated: pca, harmony, umap
```

```r
# How can I calculate the average expression of all cells within a cluster?
cluster.averages <- AverageExpression(pbmc)
```

```
## Finished averaging RNA for cluster 0
```

```
## Finished averaging RNA for cluster 1
```

```
## Finished averaging RNA for cluster 2
```

```
## Finished averaging RNA for cluster 3
```

```
## Finished averaging RNA for cluster 4
```

```
## Finished averaging RNA for cluster 5
```

```
## Finished averaging RNA for cluster 6
```

```
## Finished averaging RNA for cluster 7
```

```
## Finished averaging RNA for cluster 8

## Finished averaging RNA for cluster 9

## Finished averaging RNA for cluster 10

## Finished averaging RNA for cluster 11

## Finished averaging RNA for cluster 12

## Finished averaging RNA for cluster 13

## Finished averaging RNA for cluster 14

## Finished averaging RNA for cluster 15

## Finished averaging SCT for cluster 0

## Finished averaging SCT for cluster 1

## Finished averaging SCT for cluster 2

## Finished averaging SCT for cluster 3

## Finished averaging SCT for cluster 4

## Finished averaging SCT for cluster 5

## Finished averaging SCT for cluster 6

## Finished averaging SCT for cluster 7

## Finished averaging SCT for cluster 8

## Finished averaging SCT for cluster 9

## Finished averaging SCT for cluster 10

## Finished averaging SCT for cluster 11

## Finished averaging SCT for cluster 12

## Finished averaging SCT for cluster 13

## Finished averaging SCT for cluster 14

## Finished averaging SCT for cluster 15
```

```r
head(cluster.averages[["RNA"]][, 1:5])
```

```
##                       0           1           2          3           4
## AL627309.1 0.000000000 0.002389822 0.002541837 0.00000000 0.000000000
## AL669831.5 0.015263925 0.023898217 0.040669392 0.03124149 0.003602268
## FAM87B     0.000000000 0.007169465 0.002541837 0.00000000 0.000000000
## LINC00115  0.017444486 0.028677861 0.052662496 0.04828230 0.010806804
## FAM41C     0.002180561 0.011949109 0.002541837 0.00000000 0.000000000
## AL645608.3 0.004361121 0.028004605 0.080622703 0.02556122 0.000000000
```

```r
# Return this information as a Seurat object (enables downstream plotting and analysis) First,
# replace spaces with underscores '_' so ggplot2 doesn't fail
orig.levels <- levels(pbmc)
Idents(pbmc) <- gsub(pattern = " ", replacement = "_", x = Idents(pbmc))
orig.levels <- gsub(pattern = " ", replacement = "_", x = orig.levels)
levels(pbmc) <- orig.levels
cluster.averages <- AverageExpression(pbmc, return.seurat = TRUE)
```

```
## Finished averaging RNA for cluster 0
```

```
## Finished averaging RNA for cluster 1
```

```
## Finished averaging RNA for cluster 2
```

```
## Finished averaging RNA for cluster 3
```

```
## Finished averaging RNA for cluster 4
```

```
## Finished averaging RNA for cluster 5
```

```
## Finished averaging RNA for cluster 6
```

```
## Finished averaging RNA for cluster 7
```

```
## Finished averaging RNA for cluster 8
```

```
## Finished averaging RNA for cluster 9
```

```
## Finished averaging RNA for cluster 10
```

```
## Finished averaging RNA for cluster 11
```

```
## Finished averaging RNA for cluster 12
```

```
## Finished averaging RNA for cluster 13
```

```
## Finished averaging RNA for cluster 14
```

```
## Finished averaging RNA for cluster 15

## Finished averaging SCT for cluster 0

## Finished averaging SCT for cluster 1

## Finished averaging SCT for cluster 2

## Finished averaging SCT for cluster 3

## Finished averaging SCT for cluster 4

## Finished averaging SCT for cluster 5

## Finished averaging SCT for cluster 6

## Finished averaging SCT for cluster 7

## Finished averaging SCT for cluster 8

## Finished averaging SCT for cluster 9

## Finished averaging SCT for cluster 10

## Finished averaging SCT for cluster 11

## Finished averaging SCT for cluster 12

## Finished averaging SCT for cluster 13

## Finished averaging SCT for cluster 14

## Finished averaging SCT for cluster 15

## Centering and scaling data matrix
```
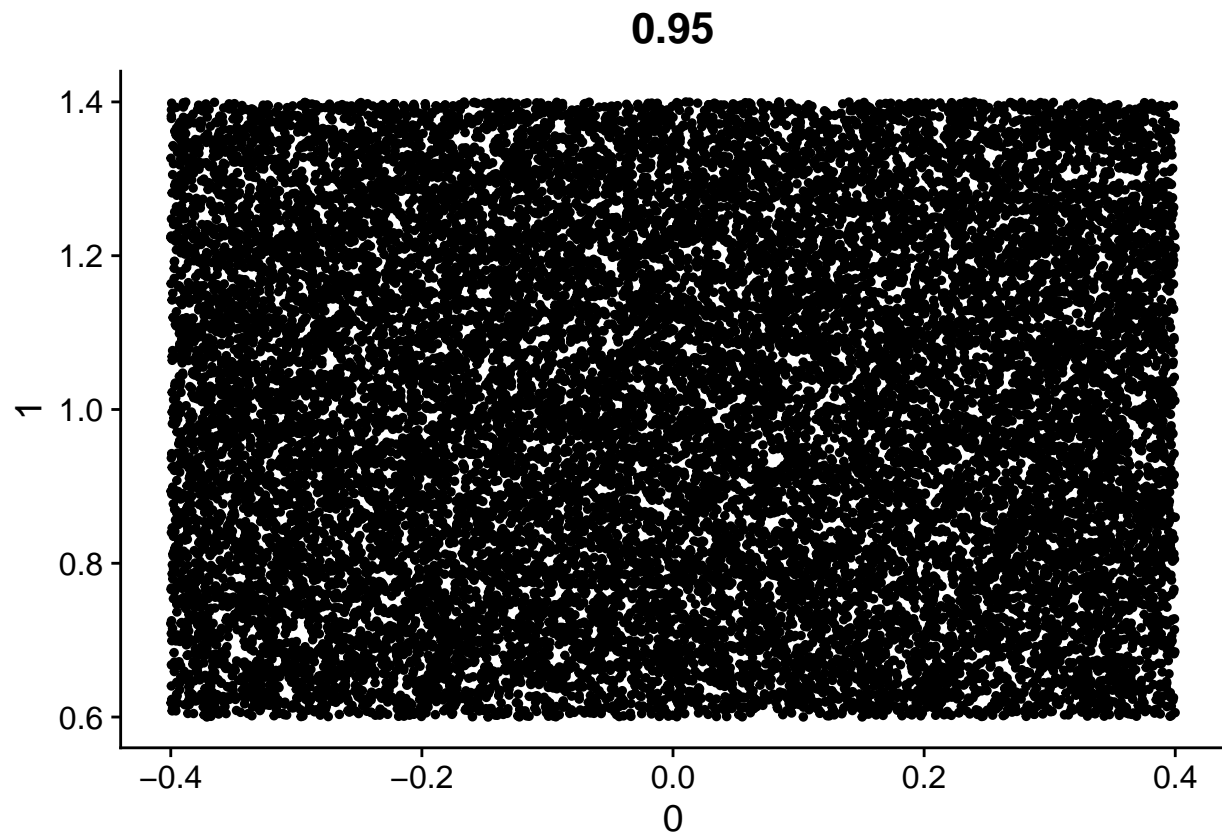
```
cluster.averages
```

```
## An object of class Seurat
## 43718 features across 16 samples within 2 assays
## Active assay: SCT (21043 features, 0 variable features)
##  1 other assay present: RNA
```

```
# How can I plot the average expression of X cells vs. Y cells?  Pass do.hover = T for an
# interactive plot to identify gene outliers
CellScatter(cluster.averages, cell1 = "0", cell2 = "1")
```

0.95

```
# How can I calculate expression averages separately for each replicate?
#cluster.averages <- AverageExpression(pbmc, return.seurat = TRUE, add.ident = "replicate")
#CellScatter(cluster.averages, cell1 = "cluster 1_rep1", cell2 = "cluster 1_rep2")
```