

## Dimension reduction

```
library(Seurat)
library(ggplot2)
library(reshape2)
library(plyr)
library(RColorBrewer)
```

```
subdataset <- readRDS('../tmp/crc_smc.malignantcells.Rds')
head(subdataset@meta.data, n=3); nrow(subdataset@meta.data)
```

```
##                                nCount_RNA nFeature_RNA      Library
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1      35998       4823 PM-PS-0001-T-A1
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1      31383       5252 PM-PS-0001-T-A1
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1       7302       1713 PM-PS-0001-T-A1
##                                Patient  Sample Cell_subtype RNA_snn_res.0.3
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1  SMC01  SMC01-T         CMS2             2
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1  SMC01  SMC01-T         CMS2             0
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1  SMC01  SMC01-T         CMS2             2
```

```
## [1] 17334
```

```
pvals <- data.frame(subdataset@reductions$pca@jackstraw$overall.p.values)
pcs_use <- pvals[pvals$Score > 0.001, 'PC'][1]-1; pcs_use # 28
```

```
## [1] 28
```

```
subdataset <- RunTSNE(subdataset, dims=1:pcs_use, reduction = "pca", reduction.key='tSNE', dim.embed=2,
```

```
## Warning: All keys should be one or more alphanumeric characters followed by an
## underscore '_', setting key to tSNE_
```

```
subdataset <- RunUMAP(subdataset, dims=1:pcs_use, reduction = "pca", reduction.key='UMAP', n.components=
```

```
## Warning: The default method for RunUMAP has changed from calling Python UMAP via reticulate to the R
## To use Python UMAP via reticulate, set umap.method to 'umap-learn' and metric to 'correlation'
## This message will be shown once per session
```

```
## 12:43:10 UMAP embedding parameters a = 1.262 b = 1.003
```

```
## 12:43:10 Read 17334 rows and found 28 numeric columns
```

```
## 12:43:10 Using Annoy for neighbor search, n_neighbors = 30
```

```
## 12:43:10 Building Annoy index with metric = cosine, n_trees = 50

## 0%   10   20   30   40   50   60   70   80   90  100%

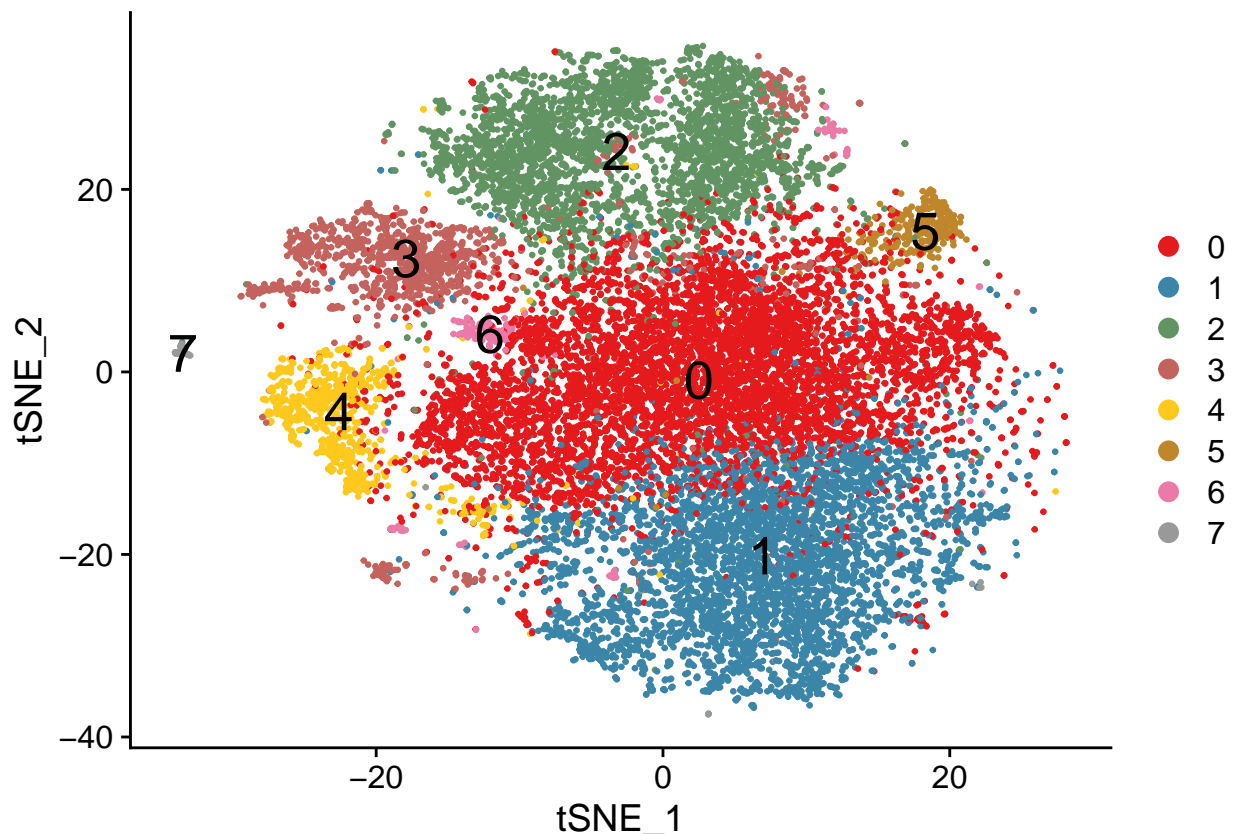
## [----|----|----|----|----|----|----|----|----|

## *****|
## 12:43:13 Writing NN index file to temp file /var/folders/sd/2bnf84cj77b2wtx4z2174nk40000gn/T//RtmpRO
## 12:43:13 Searching Annoy index using 1 thread, search_k = 3000
## 12:43:21 Annoy recall = 100%
## 12:43:22 Commencing smooth kNN distance calibration using 1 thread
## 12:43:23 Initializing from normalized Laplacian + noise
## 12:43:23 Commencing optimization for 200 epochs, with 762208 positive edges
## 12:43:37 Optimization finished

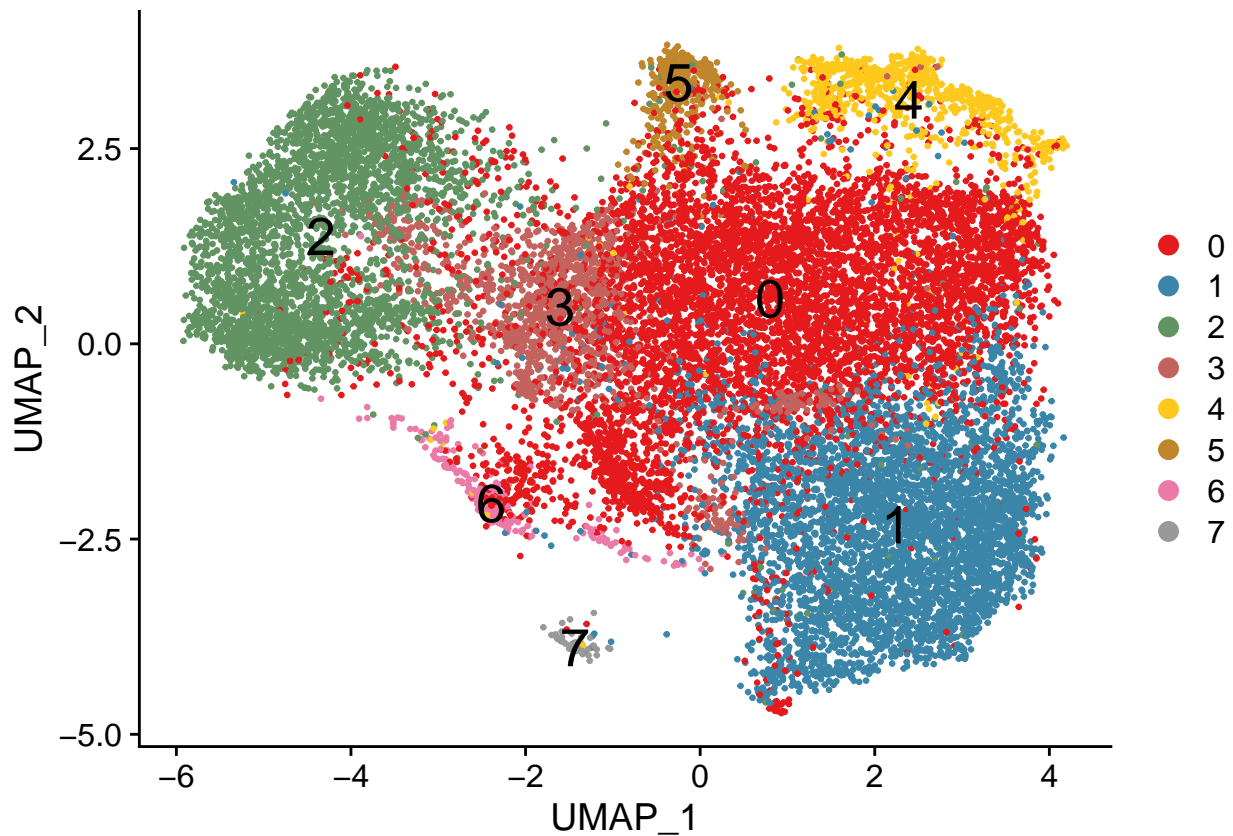
## Warning: All keys should be one or more alphanumeric characters followed by an
## underscore '_', setting key to UMAP_

getPalette <- colorRampPalette(brewer.pal(9, 'Set1'))
DimPlot(subdataset, reduction = 'tsne', pt.size = .5, label = T, label.size = 7, cols = getPalette(8))

## Warning: Using 'as.character()' on a quosure is deprecated as of rlang 0.3.0.
## Please use 'as_label()' or 'as_name()' instead.
## This warning is displayed once per session.
```



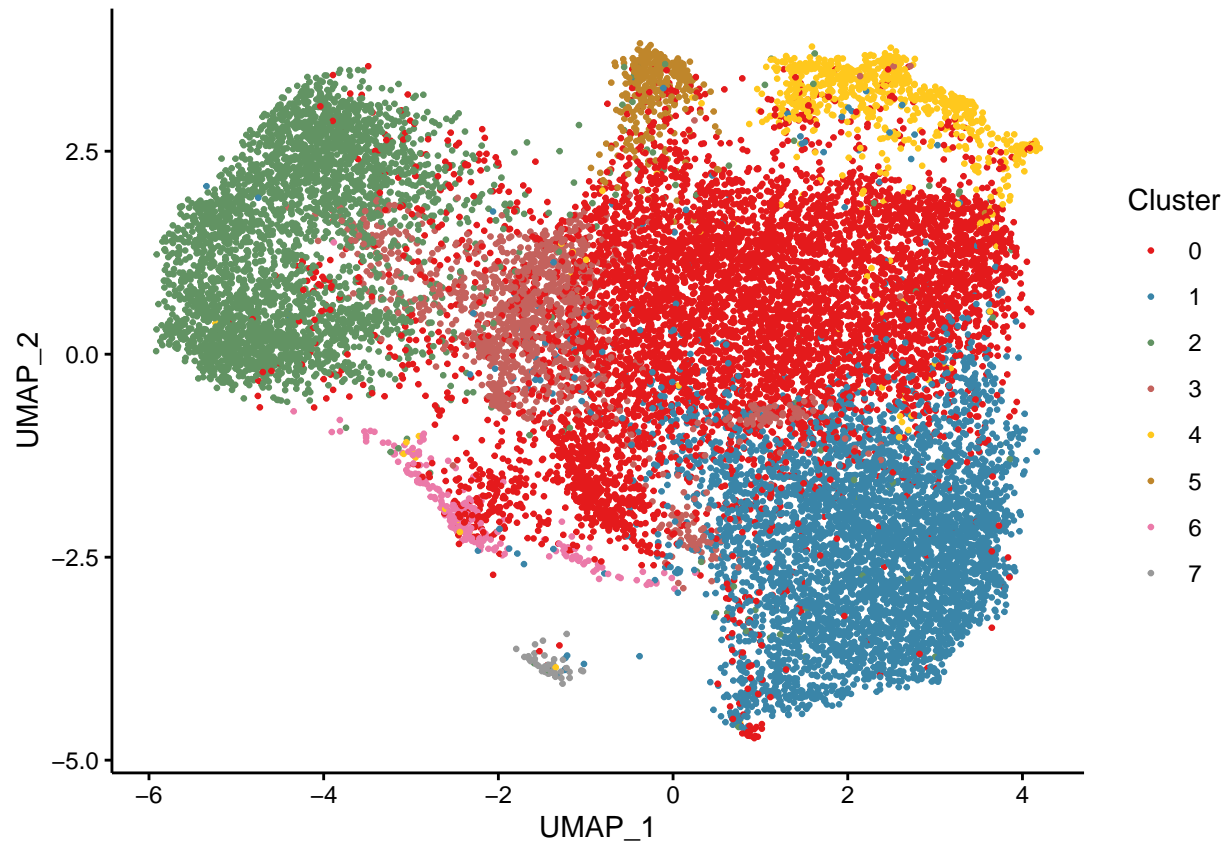
```
DimPlot(subdataset, reduction = 'umap', pt.size = .5, label = T, label.size = 7, cols = getPalette(8))
```



```
umap_coord <- data.frame(as.matrix(Embeddings(subdataset, reduction = 'umap')), check.names = F, check.names = F)
umap_coord$Cluster <- subdataset@meta.data$RNA_snn_res.0.3
umap_coord$Cluster <- factor(umap_coord$Cluster, levels = as.character(c(0:7)))
head(umap_coord)
```

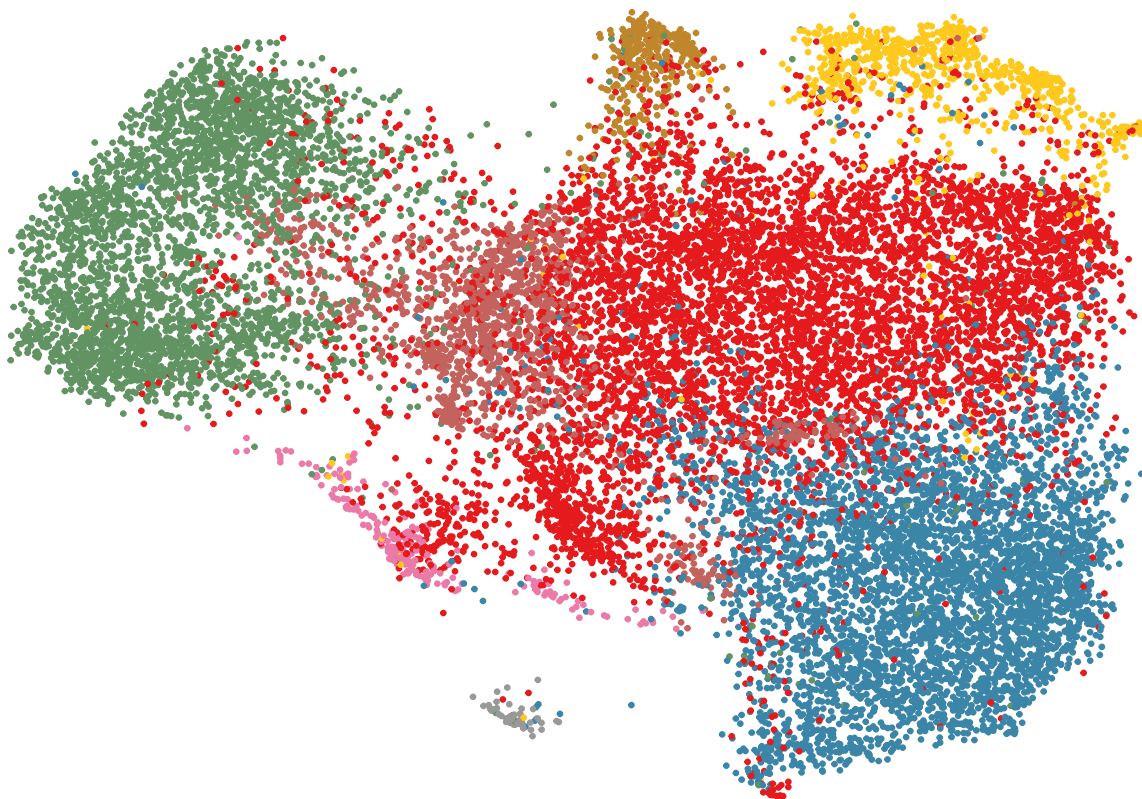
##		UMAP_1	UMAP_2	Cluster
##	AAACCTGCATACGCCG-1-PM-PS-0001-T-A1	-5.4052503	0.10750169	2
##	AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1	1.3099583	0.82416450	0
##	AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1	-3.2374978	2.23553752	2
##	AAACGGGAGGGAAACA-1-PM-PS-0001-T-A1	0.6270821	-2.51326085	1
##	AAACGGGGTATAGGTA-1-PM-PS-0001-T-A1	3.7078457	-2.92596480	1
##	AAAGATGAGCCGAAT-1-PM-PS-0001-T-A1	-3.0437917	0.08783855	2

```
ggplot(umap_coord, aes(UMAP_1, UMAP_2, col = Cluster)) +
  geom_point(size = 0.5) +
  scale_color_manual(values = getPalette(8)) +
  theme_classic() +
  theme(panel.grid = element_blank(),
        axis.text = element_text(colour = 'black'),
        axis.ticks = element_line(colour = 'black'))
```



```
#ggsave('../res_various/res_0.3/umap.res_0.3.newHvgs.pdf', units = 'cm', width = 10, height = 8)

ggplot(umap_coord, aes(UMAP_1, UMAP_2, col = Cluster)) +
  geom_point(size = 0.5) +
  scale_color_manual(values = getPalette(8)) +
  theme_void() +
  theme(legend.position = 'none')
```



```
#ggsave('../res_various/res_0.3/umap.res_0.3.newHvgs.png', units = 'cm', width = 6, height = 6)
```