

## HW4\_EKazantseva

Preprocessing: filtering out bad cells and normalization UMAP + clustering Marker selection for clusters  
GSM4407907

### R Markdown

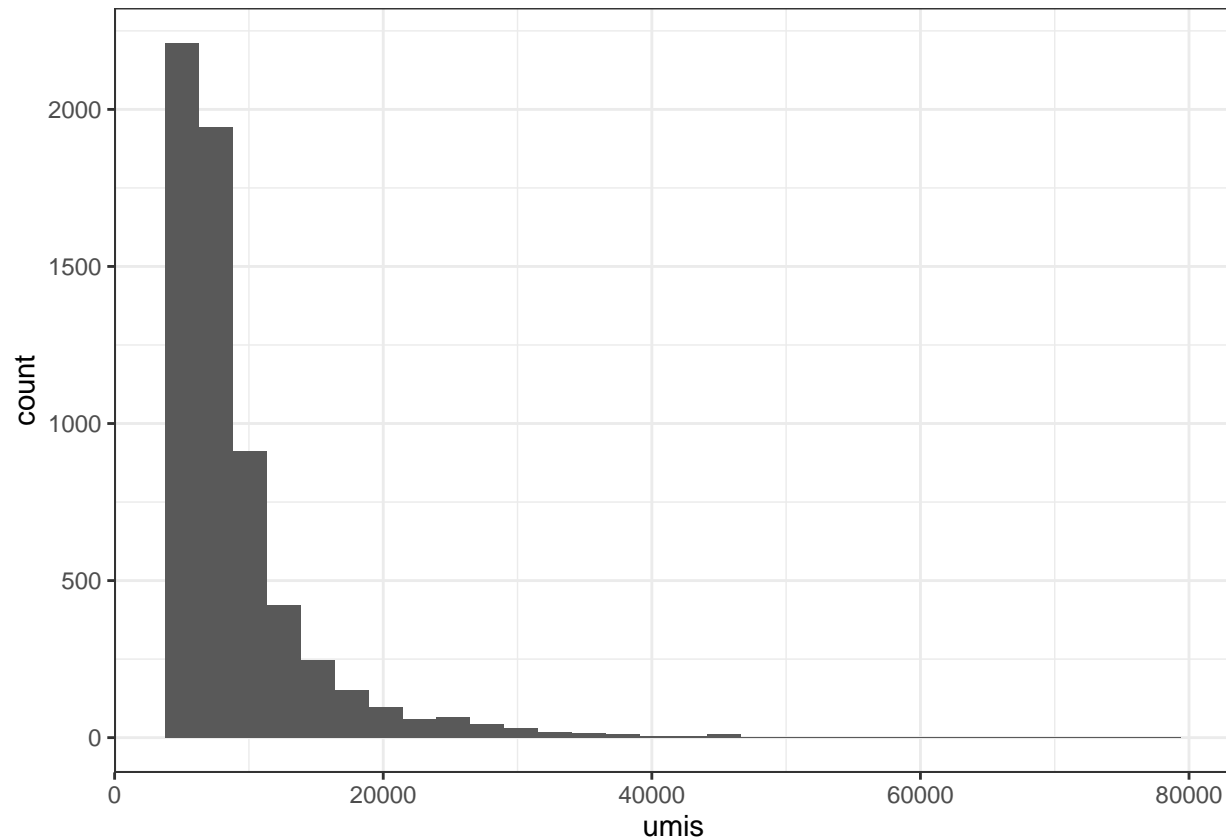
```
data <- Read10X("~/Downloads/HW_4/")  
dim(data)
```

```
## [1] 48816 6260
```

```
plotData <- data.frame(  
  umis <- colSums(data)  
)
```

```
ggplot(data=plotData, aes(x=umis)) +  
  geom_histogram() + theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



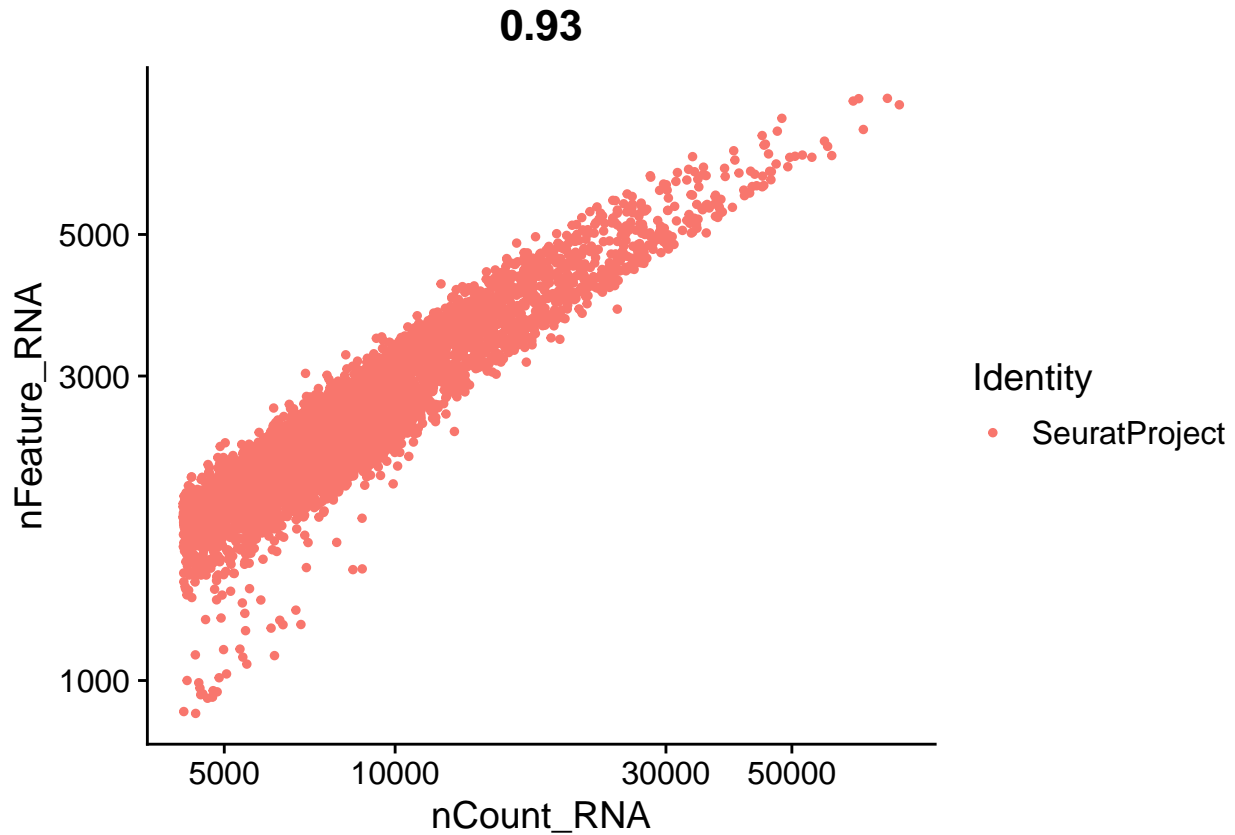
```
seurat <- CreateSeuratObject(data, min.cells = 10, min.features = 10)
```

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

```
dim(seurat)
```

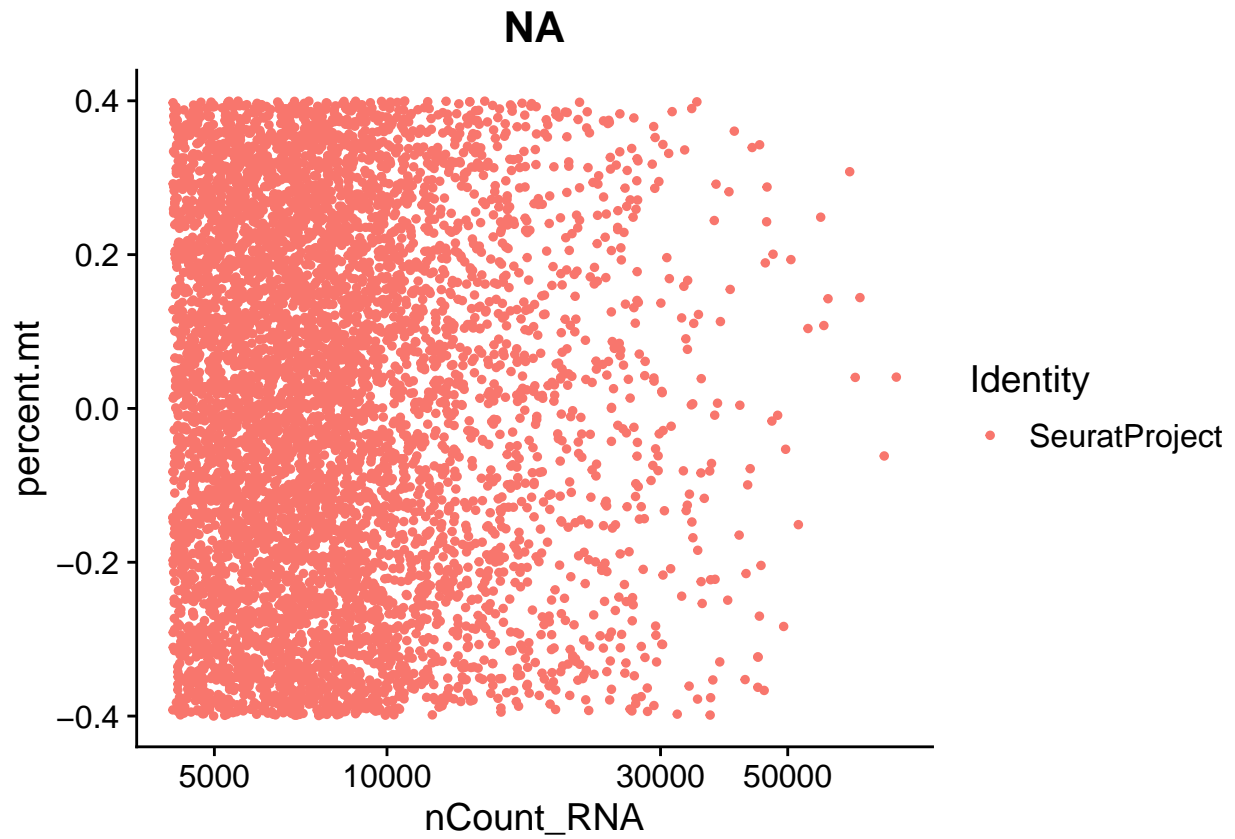
```
## [1] 18021 6260
```

```
seurat[["percent.mt"]] <- PercentageFeatureSet(seurat, pattern = "^MT-")
FeatureScatter(seurat, "nCount_RNA", "nFeature_RNA") + scale_x_log10() + scale_y_log10()
```



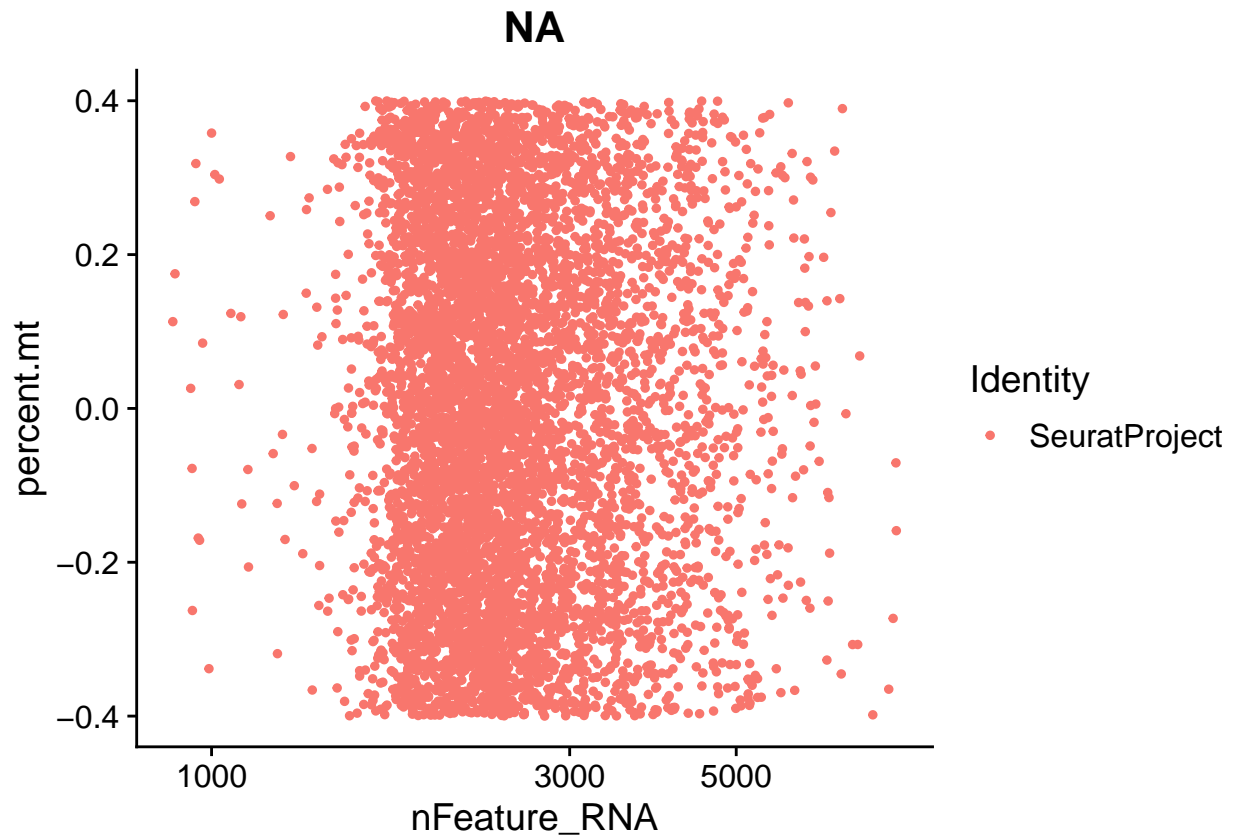
```
FeatureScatter(seurat, "nCount_RNA", "percent.mt") + scale_x_log10()
```

```
## Warning in cor(x = data[, 1], y = data[, 2]): the standard deviation is zero
```



```
FeatureScatter(seurat, "nFeature_RNA", "percent.mt") + scale_x_log10()
```

```
## Warning in cor(x = data[, 1], y = data[, 2]): the standard deviation is zero
```



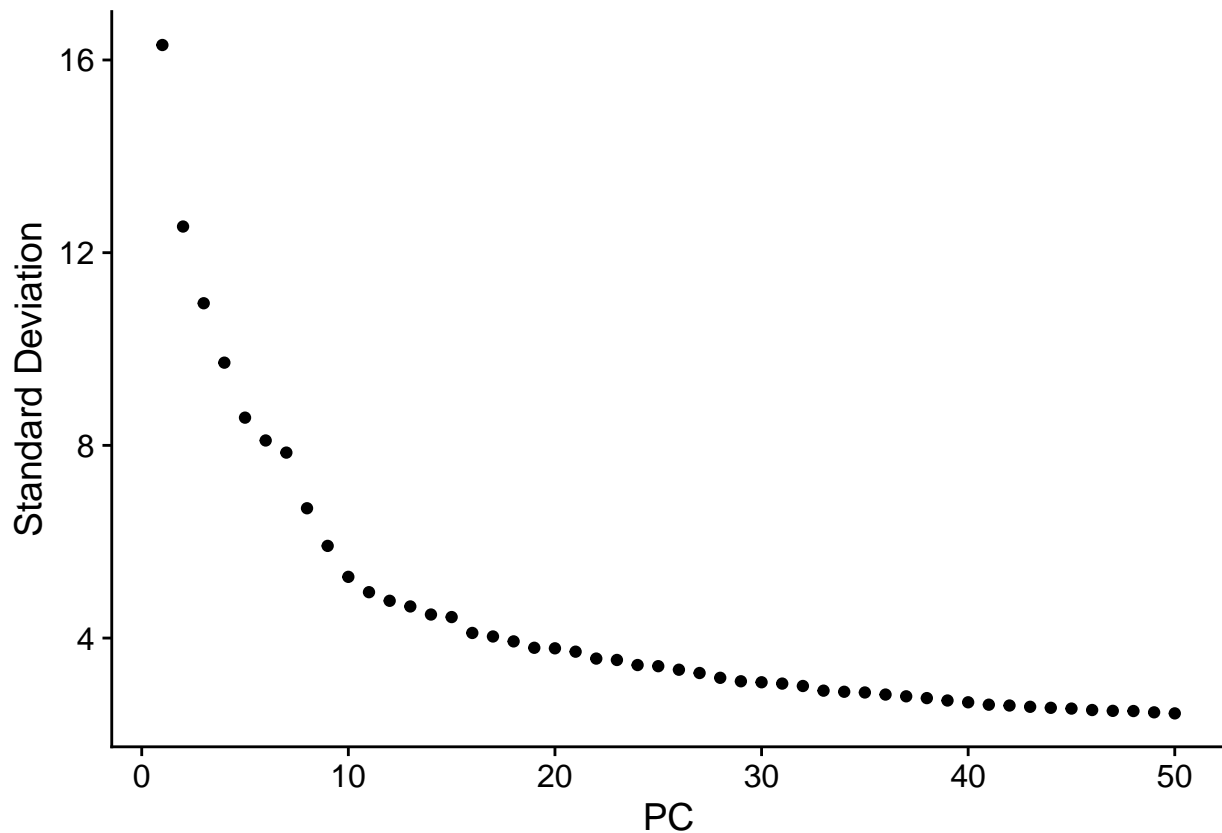
```
seurat <- subset(seurat, subset = nFeature_RNA > 1800 & percent.mt < 25)
dim(seurat)

## [1] 18021 5767

seurat <- SCTransform(seurat, vars.to.regress = "percent.mt", verbose = FALSE)

?seurat()

seurat <- RunPCA(seurat, verbose = FALSE)
ElbowPlot(seurat, ndims = 50)
```



```
seurat <- RunUMAP(seurat, dims=1:20)
```

```
## 23:36:26 UMAP embedding parameters a = 0.9922 b = 1.112
```

```
## 23:36:26 Read 5767 rows and found 20 numeric columns
```

```
## 23:36:26 Using Annoy for neighbor search, n_neighbors = 30
```

```
## 23:36:26 Building Annoy index with metric = cosine, n_trees = 50
```

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

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

```
## *****|
```

```
## 23:36:27 Writing NN index file to temp file /var/folders/n8/5xz38c314pv2_czs57hmysk4n0qqwd/T//Rtmp1f
```

```
## 23:36:27 Searching Annoy index using 1 thread, search_k = 3000
```

```
## 23:36:28 Annoy recall = 100%
```

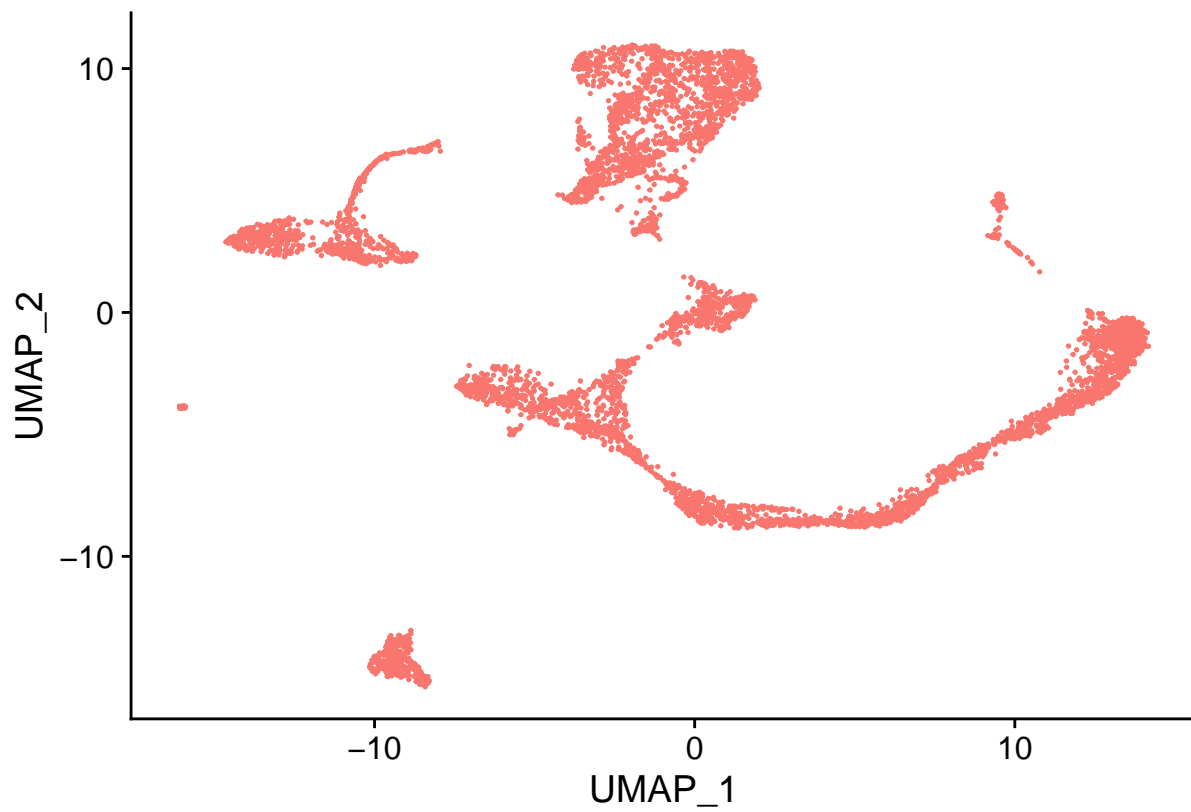
```
## 23:36:29 Commencing smooth kNN distance calibration using 1 thread
```

```
## 23:36:30 Initializing from normalized Laplacian + noise
```

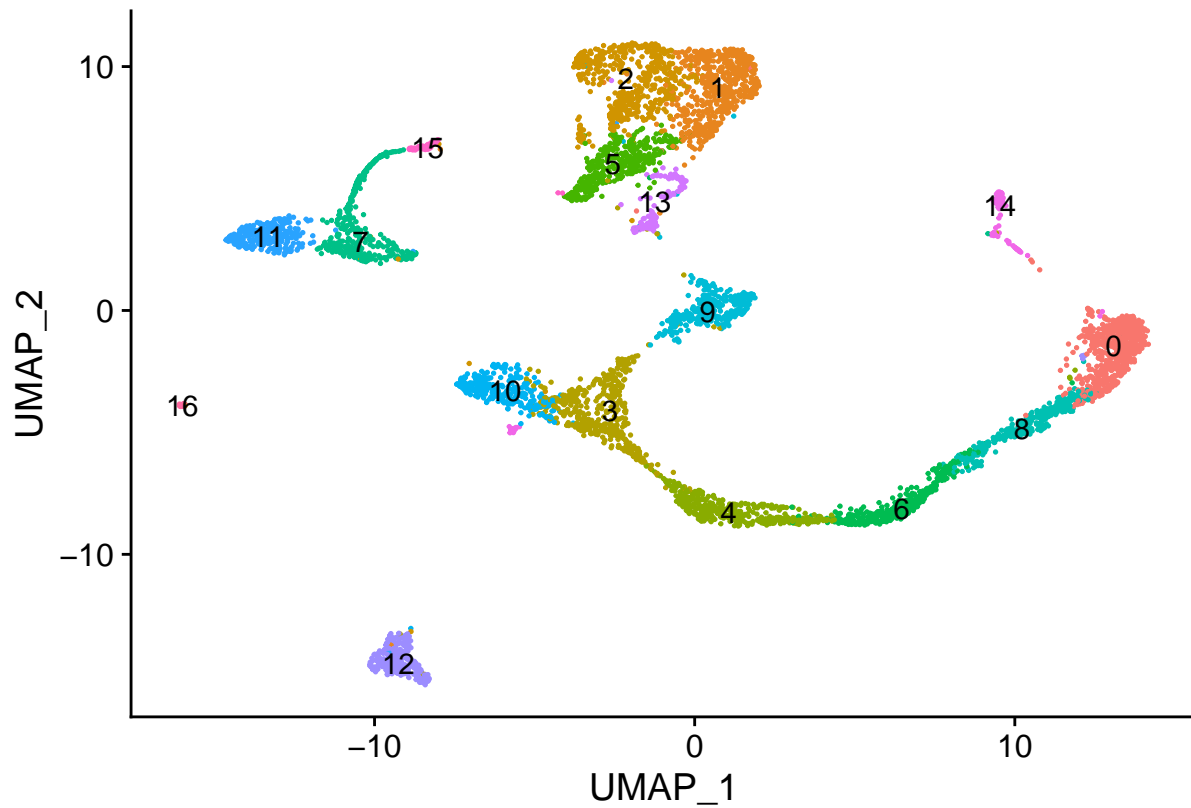
```
## 23:36:30 Commencing optimization for 500 epochs, with 240944 positive edges
```

```
## 23:36:39 Optimization finished
```

```
DimPlot(seurat, reduction = "umap") + NoLegend()
```



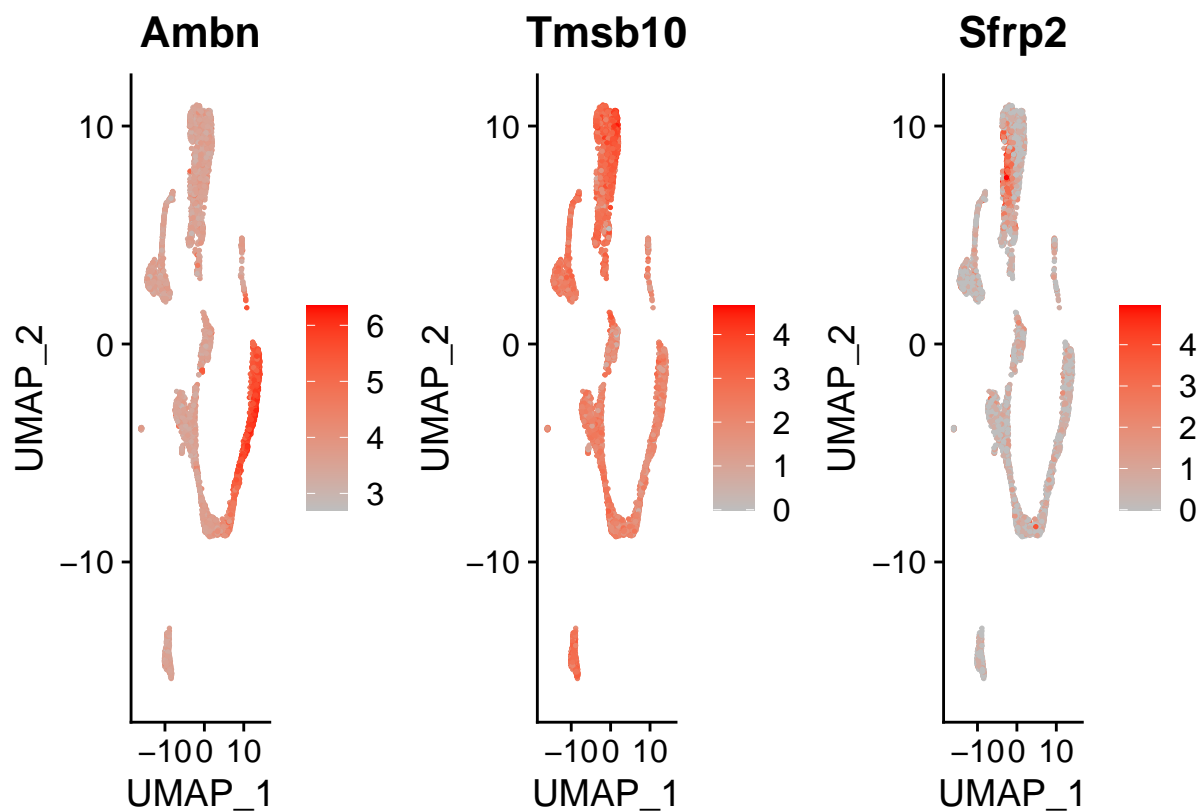
```
seurat <- FindNeighbors(seurat, dims = 1:20, verbose = FALSE)
seurat <- FindClusters(seurat, resolution=0.6, verbose = FALSE)
DimPlot(seurat, reduction = "umap", label = TRUE) + NoLegend()
```



```
allMarkers <- FindAllMarkers(seurat, max.cells.per.ident = 100, test.use = "MAST", only.pos = T)
goodMarkers <- allMarkers %>% group_by(cluster) %>% top_n(n = 1, wt = avg_log2FC) %>% pull(gene)
goodMarkers
```

```
## [1] "Ambn" "Tmsb10" "Sfrp2" "Krt17" "Dspp" "Spp1" "Clu" "Plac8"
## [9] "Clu" "Tfrc" "Hmgb2" "Dcn" "Igfbp7" "Rgs5" "ApoE" "Bglap"
## [17] "Mpz"
```

```
FeaturePlot(seurat, goodMarkers[1:3], cols=c("grey", "red"), reduction="umap", ncol=3)
```



```
VlnPlot(seurat, goodMarkers[1:3], pt.size = 0.1)
```

