scRNA-seq analisys Kazantseva

Preprocessing: Filtering out bad cells and normalization UMAP + clustering Marker selection for clusters GSM3215435

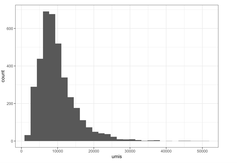
R Markdown

```
data <- Read10X('Downloads/GSM3215435/')
dim(data)
```

```
## [1] 27998  3781
```

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

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

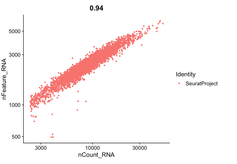
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




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

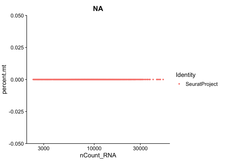
```
## [1] 13105  3781
```

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



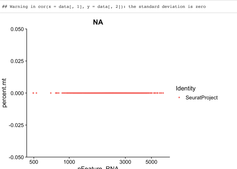

```
FeatureScatter(seurat, "nCount_RNA", "percent.mt")+scale_x_log10()
```

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




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

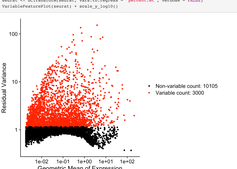
```
## Warning in cor(x = data[, 1], y = data[, 2]): the standard deviation is zero
```



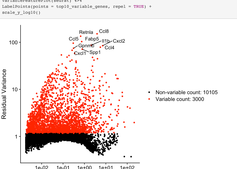

```
seurat <- subset(seurat, subset = nFeature_RNA > 1000)
dim(seurat)
```

```
## [1] 13105  3758
```

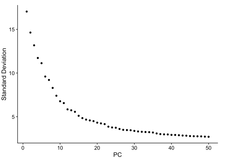
```
seurat <- SCTransform(seurat, vars.to.regress = "percent.mt", verbose = FALSE)
VariableFeaturePlot(seurat) + scale_y_log10()
```




```
top10_variable_genes <- head(VariableFeatures(seurat), 10)
VariableFeaturePlot(seurat) %>%
LabelPoints(points = top10_variable_genes, repel = TRUE) +
scale_y_log10()
```



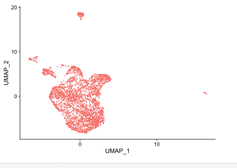

```
seurat <- RunPCA(seurat, verbose = FALSE)

ElbowPlot(seurat, ndims = 50)
```




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

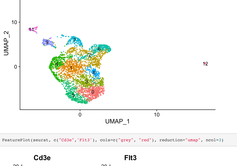
```
## Warning: The default method for RunUMAP has changed from calling Python UMAP via reticulate to the R-native UW
## This message will be shown once per session
## To use Python UMAP via reticulate, set umap.method to 'umap-learn' and metric to 'correlation'
## This message will be shown once per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

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

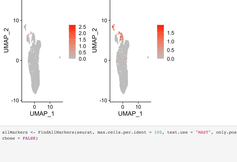



```
#seurat <- RunTSNE(seurat, dims=1:15)
#DimPlot(seurat, reduction = "tsne") + NoLegend()
```

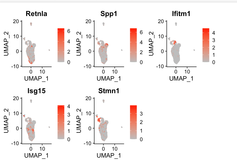
```
seurat <- FindNeighbors(seurat, dims = 1:20, verbose = FALSE)
seurat <- FindClusters(seurat, resolution=0.5, verbose = FALSE)
DimPlot(seurat, reduction = "umap", label = TRUE) + NoLegend()
```




```
FeaturePlot(seurat, c("Cd3e", "Flt3"), cols=c("grey", "red"), reduction="umap", ncol=2)
```




```
allMarkers <- FindAllMarkers(seurat, max.cells.per.ident = 100, test.use = "MAST", only.pos = T,assay = "RNA", ve
rbose = FALSE)


goodMarkers <- allMarkers %>% group_by(cluster) %>% top_n(n = 1, wt = avg_logFC) %>% pull(gene)

FeaturePlot(seurat, goodMarkers[4:12], cols=c("grey", "red"), reduction="umap", ncol=3)
```




```
VlnPlot(seurat, goodMarkers[4:12], pt.size = 0.1)
```