Single Cell Sequencing Analysis

Divya Agrawal

November 2022

```
Dataset is downloaded from Broad Insitute Single cell website-
https://singlecell.broadinstitute.org/single_cell
/study/SCP1256/visium-demo-study#study-summary
Named as - Study: Visium demo study

Codes for single cell seq analysis
install.packages("Seurat")
install.packages("tidyverse")
install.packages("dplyr")
install.packages("patchwork")
install.packages("ggplot2")

library(Seurat)
library(tidyverse)
library(dplyr)
library(patchwork)
library(ggplot2)

##load the datasets##

da.data <- Read10X(data.dir ="C:/Users/Divya Agrawal/Downloads/")
da <- CreateSeuratObject(counts = da.data, min.cells = 4, min.features = 210)

##QC and filtering##

da[["percent.mt"]] <- PercentageFeatureSet(da, pattern = "^MT-")
plot1<-FeatureScatter(da, feature1 = "nCount_RNA", feature2 = "nFeature_RNA")
plot1
plot2 <-FeatureScatter(da, feature1 = "nCount_RNA", feature2 = "percent.mt")
plot2
plot1 +plot2
da <- subset(da, subset = nFeature_RNA >215 & nFeature_RNA > 2500 & percent.mt <5)

##normalise the data##
```

```
da <- NormalizeData(da, normalization.method = "LogNormalize", scale.factor = 10000)

##Find variable Features##

da <-FindVariableFeatures(da, selection.method = "vst", mfeatures=2000)
tp10<- head(VariableFeatures(da), 10)
tp10
plot1<- VariableFeaturePlot(da)

##scale the data##

all.genes <-rownames(da)
pre_scaling <-da
da <- ScaleData(da, features = all.genes)


##run linear dimensionality reduction ##

da<-RunPCA(da, features = VariableFeatures(object = da))
print(da[["pca"]], dims = 1:5, nfeatures = 5)
VizDimLoadings(da, dims = 1:2, reduction= "pca")
DimHeatmap(da, dims = 1, cells = 500)
DimHeatmap(da, dims = 1:15, cells = 500)
da <- JackStraw(da, num.replicate = 100)
JackStrawPlot(da, dims = 1:20)
da <-ScoreJackStraw(da, dims = 1:20)
da <-ScoreJackStraw(da, dims = 1:20)
JackStrawPlot(da, dims = 1:20)


##cluster##

da <- FindNeighbors(da, dims = 1:10)
da <-FindClusters(da, resolution = 0.5)
head(Idents(da),5)
##run non linear dimensionality reduction on top of dimensionality reduction
da <- RunUMAP(da, dims = 1:10)
DimPlot(da, reduction = "umap")

##assign the biological meaning to these clusters

da.markers <-FindAllMarkers(da, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25)
da.markers %>% group_by(cluster) %>% slice_max(n=2, order_by =avg_log2FC)
da.markers
```

```
FeaturePlot(da, features = c("CTSH","CCL5","ENG","CD79A"))


##talk to a biologist

new.cluster.ids <- c("Naive CD T", "CD14+ Mono", "Memory CD4 T", "B","CD 8 T",
"FCG3A+ Mono","NK cells", "DC" ,"platelet", "MAC complex")
names(new.cluster.ids) <- levels(da)
da <- RenameIdents(da, new.cluster.ids)
DimPlot(da, reduction = "umap", label = TRUE, pt.size = 0.5) + NoLegend()
```

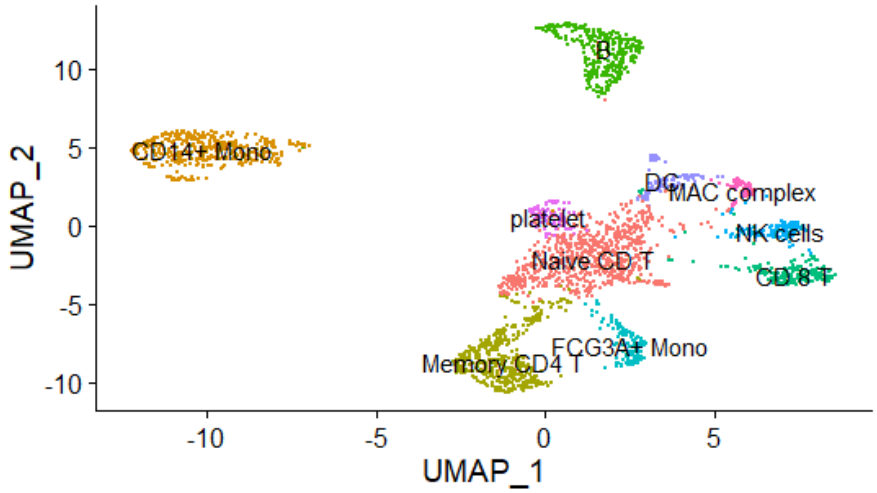

Figure 1: UMAP Plot