Standard Workflow for the Single Cell RNA Seq

1. Count Matrix
2. QC and filtering
3. Normalisation
4. Linear dimensionality reduction (PCA)
5. Scale data
6. Identify highly variable genes
7. Clustering
8. Non-linear dimensionality reduction (UMAP/t-SNE)

Downstream analysis

* Cluster identification
* Perform differential gene expression between clusters/differential chromatin accessibility analysis
* Inferring trajectories/lineage

R packages used are-

* Seurat
* Tidyverse

Codes-

Download the datasets from the Broad Institute Single cell Portal and Donwload matrix gene expression file, features and barcodes file in tsv.gz format always.

##codes for single cell seq analysis###

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()

