#安装导入包：

install.packages("pacman") # pacman包如果已经安装则不需要这句

pacman::p\_load(hdf5r, dplyr, Seurat, patchwork, SingleR, celldex, pheatmap, ggplot2, cowplot,harmony)

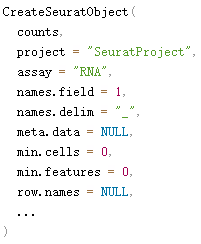
#读取文件

data\_sample <- Read10X\_h5("C:/Users/pp/Desktop/新建文件夹/GSM6177599\_NYU\_BRCA0\_Vis\_processed\_filtered\_feature\_bc\_matrix.h5") #h5,正常Read10X

#创建Seurat对象

data\_seurat <- CreateSeuratObject(data\_sample,project = "data\_sample", min.cells = 3, min.features = 200)

CreateSeuratObject参数：



names.field: 对于每个细胞的初始标识类，从细胞名字识别字段。

E.g. If your cells are named as BARCODE\_CLUSTER\_CELLTYPE in the input matrix, set names.field to 3 to set the initial identities to CELLTYPE.

names.delim: 从细胞的列名中选择分隔符。

E.g. If your cells are named as BARCODE-CLUSTER-CELLTYPE, set this to “-” to separate the cell name into its component parts for picking the relevant field.

meta.data: 要添加到修拉对象的其他单元格级元数据。应该是一个 data.frame，其中行是单元格名称，列是附加元数据字段。元数据中的行名需要与计数矩阵的列名匹配。

数据集中测到的少于200个基因的细胞（min.features = 200）和少于3个细胞覆盖的基因（min.cells = 3）被过滤掉

#质控

#计算线粒体占比

data\_seurat[["percent.mt"]] <- PercentageFeatureSet(data\_seurat, pattern = "^MT-")

#使用小提琴图可视化QC指标#nFeature\_RNA代表每个细胞测到的基因数目，nCount代表每个细胞测到所有基因的表达量之和，percent.mt代表测到的线粒体基因的比例。

VlnPlot(data\_seurat, features = c("nFeature\_RNA", "nCount\_RNA", "percent.mt"), ncol = 3)

#去除线粒体基因表达比例过高的细胞，和一些极值细胞。

data\_seurat <- subset(data\_seurat, subset = nFeature\_RNA > 200 & nFeature\_RNA < 2500 & percent.mt < 5)

subset函数：从某一个数据框中选择出符合某条件的数据或是相关的列。

##标准化 全局缩放归一化方法“lognormalize”,结果log转换

data\_seurat <- NormalizeData(data\_seurat, normalization.method = "LogNormalize", scale.factor = 10000)