

# QC CM1\_2

```
library(Seurat)
library(ggplot2)
```

load SoupX adjusted count matrix to Seurat

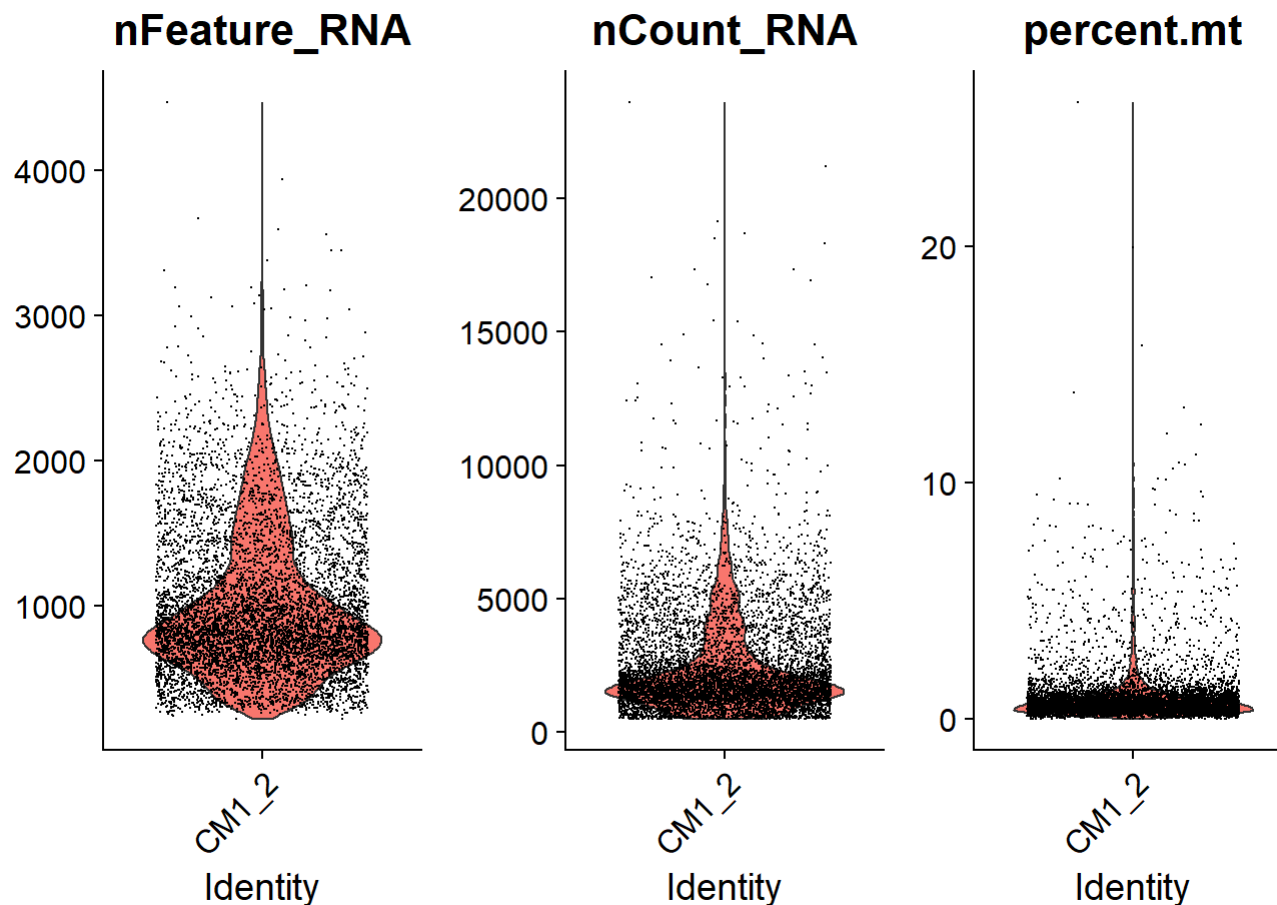
```
CM1_2=Read10X('G://covid19scRNAseq/CM1_2/outs/desoup5')
CM1_2=CreateSeuratObject(counts = CM1_2, project = "CM1_2")
```

QC and select cells for further analysis calculate mitochondrial QC metrics

```
CM1_2[["percent.mt"]] <- PercentageFeatureSet(CM1_2, pattern = "^MT-")
```

Visualize QC metrics as a violin plot

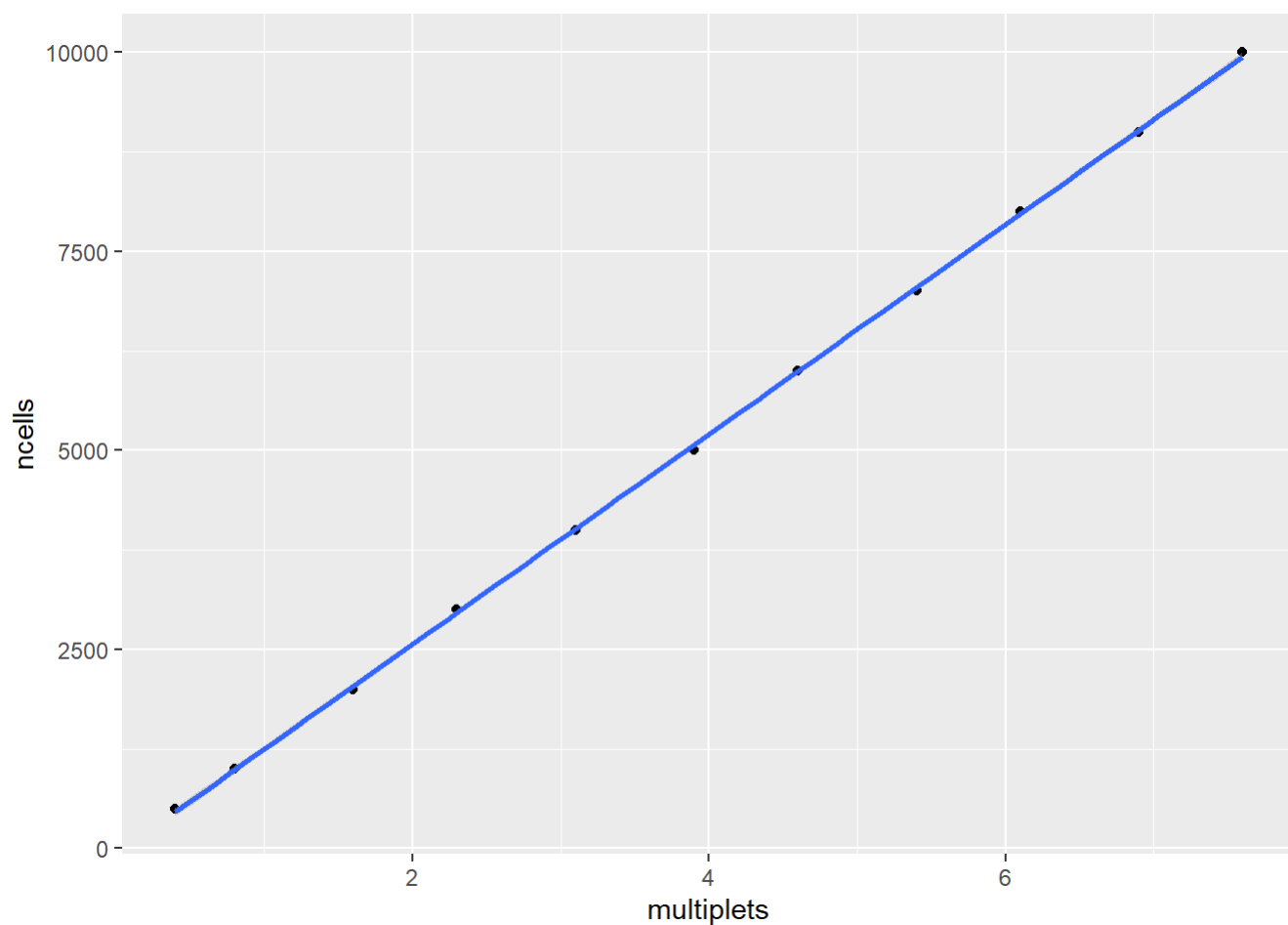
```
VlnPlot(CM1_2, features = c("nFeature_RNA", "nCount_RNA", "percent.mt"), ncol = 3, pt.size=0.0001)
```



Creat linear regression fit model using multiplet ratio predicted by 10X Genomic user manual

```
ncells <- c(500,1000,2000,3000,4000,5000,6000,7000,8000,9000,10000)
multiplets <- c(0.4,0.8,1.6,2.3,3.1,3.9,4.6,5.4,6.1,6.9,7.6)
curva <- data.frame(multiplets,ncells)
ggplot(curva, aes(multiplets, ncells)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
fit <- lm(multiplets ~ ncells, curva)

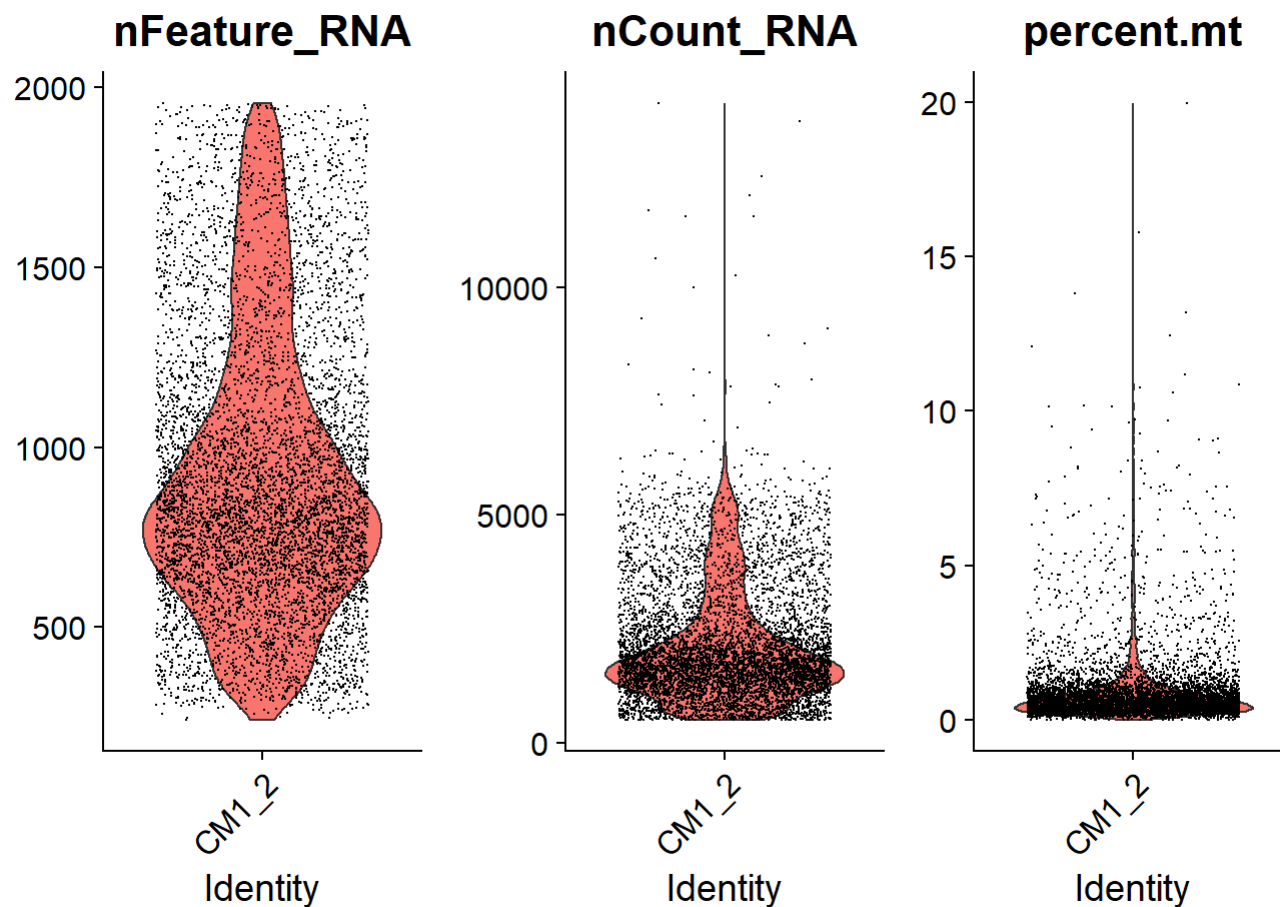
model <- function(x){
  0.0007589*x + 0.0527214
}

perc <- model(ncol(CM1_2))
q = (100 - perc)/100
feature_limit <- quantile(CM1_2$nFeature_RNA, q)
```

select cells for further analysis

```
CM1_2 <- subset(CM1_2, subset = nFeature_RNA > 200 & nCount_RNA>500 & nFeature_RNA < feature_limit & percent.mt < 20)
```

```
VlnPlot(object=CM1_2,c("nFeature_RNA", "nCount_RNA", "percent.mt"), pt.size = 0.0001)
```



Set up metadata for patient and group information, save final data

```
CM1_2$group="Moderate"  
CM1_2$patient="CM1_2"  
  
save(CM1_2, file="CM1_2.rds")
```