Introduction to single-cell transcriptomics Exercise series 1

Exercise 1 (basic QC)

Go to GEO and download the file GSE77288_molecules-raw-single-per-sample.txt.gz. It contains deep single-cell data from pluripotent stem cells obtained from three individuals.

Load the data:

```
scdat <- read.csv(
            "GSE77288_molecules-raw-single-per-sample.txt",
            sep="\t",stringsAsFactors=F)
scdat[1:5,1:10]
counts <- t(data.matrix(scdat[,-(1:3)]))
counts[1:5,1:10]
colnames(counts) <- paste(scdat$individual,scdat$replicate,scdat$well,sep=".")
counts[1:5,1:10]
counts <- counts[-grep("ERCC",rownames(counts)),] # removes spiked genes
```

Using the function `colSums()`, compute the total number of UMIs and genes per cell. Generate plots (histograms or others) and propose a policy to discard cells harboring too low complexity or excessively large UMI totals.