Covariate Analysis

Ricardo Martins-Ferreira

2023-07-24

Example for the "Study" variable

```
libs <- c("Seurat","RColorBrewer","ggplot2","SingleCellExperiment","scater")
suppressMessages(suppressWarnings(sapply(libs, require, character.only=TRUE)))
```

```
##               Seurat         RColorBrewer              ggplot2
##                 TRUE                 TRUE                 TRUE
## SingleCellExperiment               scater
##                 TRUE                 TRUE
```

The Seurat object corresponding to the total of 64,438 nuclei has been annotated based on the clinicopathological information of each subject.

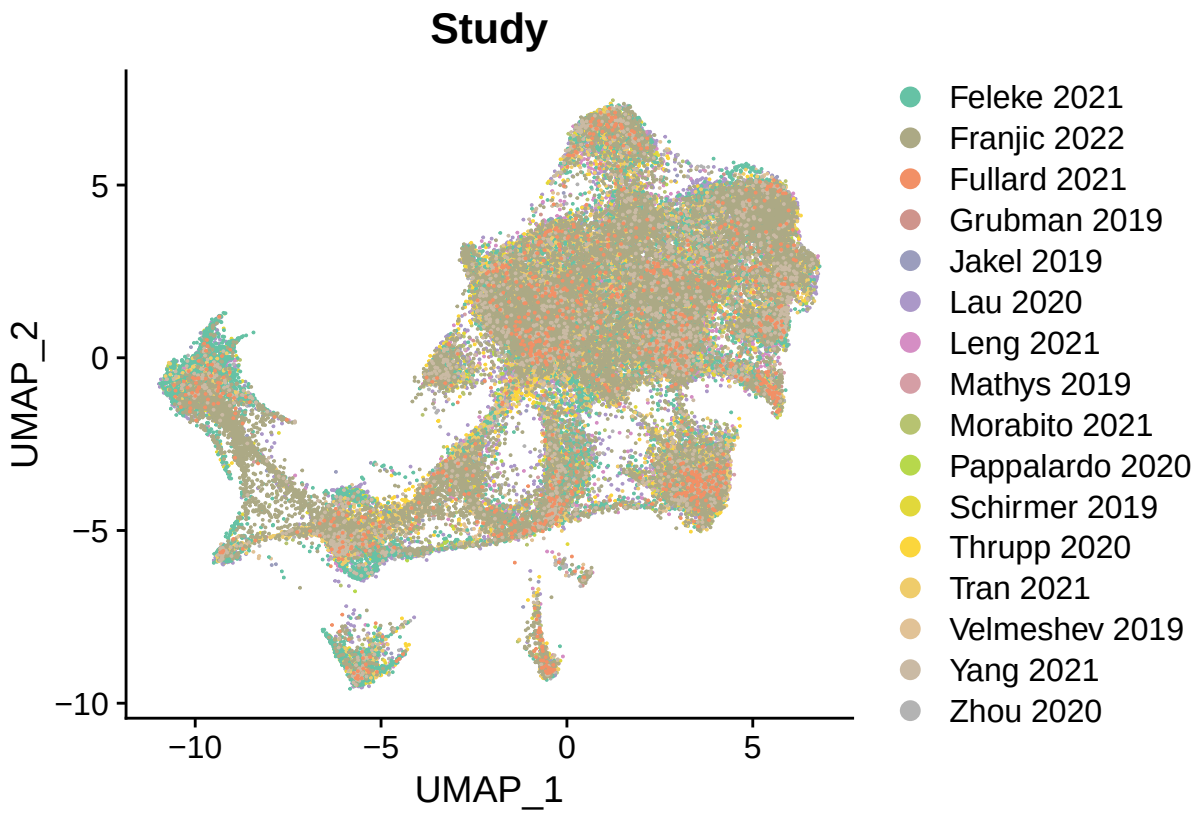
Check the distribution of the nuclei across the variable

```
table(Seurat@meta.data$Study)
```

```
##
##     Feleke 2021    Franjic 2022    Fullard 2021    Grubman 2019      Jakel 2019
##            9741           16342            1060             340            1002
##        Lau 2020       Leng 2021     Mathys 2019   Morabito 2021 Pappalardo 2020
##            6631            5432             913            3883             939
##   Schirmer 2019     Thrupp 2020       Tran 2021  Velmeshev 2019       Yang 2021
##            1480            3992            3826            3285            2003
##       Zhou 2020
##            3569
```

Visualization of the distribution in the UMAP

```
nb.cols <- 16
mycolors <- colorRampPalette(brewer.pal(8, "Set2"))(nb.cols)

DimPlot(Seurat, reduction = "umap", group.by = "Study", cols = mycolors, pt.size = 0.01)
```



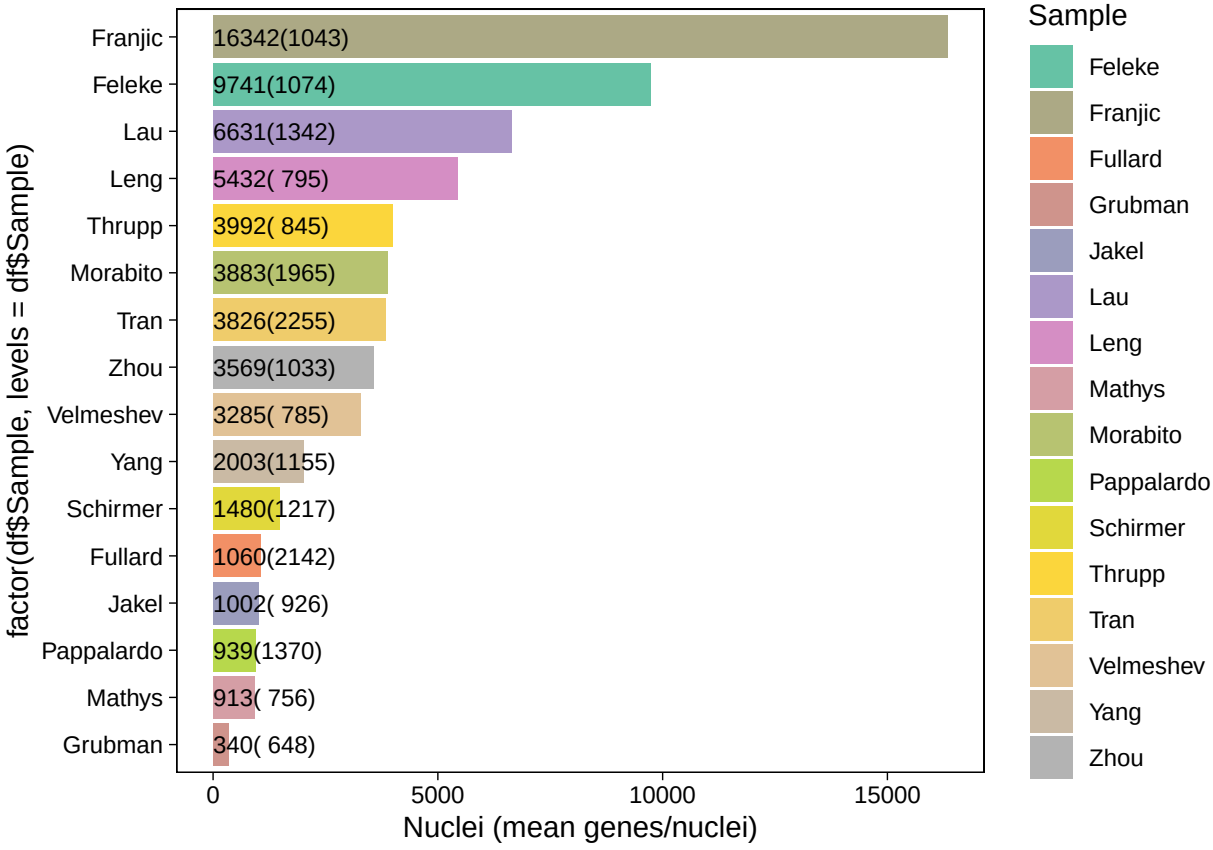
Calculate the number of nuclei and mean genes per nuclei

```
# subset the Seurat object by study
Mathys<- Seurat[,Seurat@meta.data$Study=="Mathys 2019"]
Grubman<- Seurat[,Seurat@meta.data$Study=="Grubman 2019"]
Leng<- Seurat[,Seurat@meta.data$Study=="Leng 2021"]
Morabito<- Seurat[,Seurat@meta.data$Study=="Morabito 2021"]
Lau<- Seurat[,Seurat@meta.data$Study=="Lau 2020"]
Zhou<- Seurat[,Seurat@meta.data$Study=="Zhou 2020"]
Pappalardo<- Seurat[,Seurat@meta.data$Study=="Pappalardo 2020"]
Thrupp<- Seurat[,Seurat@meta.data$Study=="Thrupp 2020"]
Jakel<- Seurat[,Seurat@meta.data$Study=="Jakel 2019"]
Schirmer<- Seurat[,Seurat@meta.data$Study=="Schirmer 2019"]
Velmeshev<- Seurat[,Seurat@meta.data$Study=="Velmeshev 2019"]
Feleke<- Seurat[,Seurat@meta.data$Study=="Feleke 2021"]
Tran<- Seurat[,Seurat@meta.data$Study=="Tran 2021"]
Franjic<- Seurat[,Seurat@meta.data$Study=="Franjic 2022"]
Yang<- Seurat[,Seurat@meta.data$Study=="Yang 2021"]
Fullard<- Seurat[,Seurat@meta.data$Study=="Fullard 2021"]

# apply a loop to calculate the number of nuclei and mean genes in each subset object
main.list <- list(Mathys, Grubman,Leng, Morabito, Lau, Zhou, Pappalardo, Thrupp, Jakel, Schirmer, Velmes

my.files <- c("Mathys", "Grubman","Leng", "Morabito", "Lau", "Zhou",
              "Pappalardo", "Thrupp", "Jakel", "Schirmer","Velmeshev",
```

```
              "Feleke", "Tran", "Franjic", "Yang", "Fullard")

df <- data.frame(Sample=my.files, Nuclei="", Genes="")
number_nuclei= numeric(length(my.files))
number_genes = numeric(length(my.files))

for (i in 1:length(main.list)) {

  number_nuclei[i]<- nrow(main.list[[i]]@meta.data)
  number_genes[i] <- mean(main.list[[i]]@meta.data$nFeature_RNA)
}

df$Nuclei = number_nuclei
df$Genes = number_genes
df$Genes = format(round(df$Genes, 0))

df$Label <- paste0(df$Nuclei,"(",df$Genes,")")
df <- df[order(df$Nuclei),]
```

Plot the number of nuclei and mean genes per nuclei

```
ggplot(df, aes(x= factor(df$Sample, levels = df$Sample), y=Nuclei, fill=Sample)) +
  geom_bar(stat="identity",position = "stack")+
  geom_text(aes(label = Label),
            hjust=0,position = position_fill(vjust = 0),
            size = 3,colour="black")+
  coord_flip()+
  ylab(label = "Nuclei (mean genes/nuclei)")+
  scale_fill_manual(values = mycolors)+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.text = element_text(colour="black"))
```



Percentage of variance explained

The percentage of variance explained is calculated with the "RNA" assay.

The Seurat object is first converted to sce.

```
# Set "RNA" as active assay, normalize and scale it
DefaultAssay(Seurat)<- "RNA"
Seurat <- NormalizeData(Seurat)
Seurat <- ScaleData(Seurat)

# convert Seurat to sce
sce <- SingleCellExperiment(assays = list(logcounts = as.matrix(Seurat@assays$RNA@counts)),
                            colData = Seurat@meta.data)
```

```
# plot percentage of variance explained
plotExplanatoryVariables(sce, variables = c("Gender","Age_conc",
                                            "Tissue_type","Group",
                                            "Study","Subject",
                                            "Tissue_condition"))
```

The output corresponds to Supplementary Figure 3G in the manuscript.