Nasal gene expression - Children/Adults/Adults-Older

Load libraries

```
if (!require("here", quietly = TRUE)) install.packages("here")
if (!require("ggpubr", quietly = TRUE)) install.packages("ggpubr")
if (!require("ggfortify", quietly = TRUE)) install.packages("ggfortify")
if (!require("RColorBrewer", quietly = TRUE)) install.packages("RColorBrewer")
if (!require("scater", quietly = TRUE)) BiocManager::install("scater")
```

Read input files

```
#Gene count files
genecounts_raw <- read.delim(here("data", "gene_counts_merged_datasets_adults_children.txt"))

genecounts_fpkm <- read.delim(here('results','FPKM_merged_datasets_children_adults_older.txt'))

genecounts_combat <- read.delim(here('results','gene_counts_batch-effect_correction_Adults-Children_UK_

genecounts_scBatch <- read.delim(here('results','gene_counts_batch-effect_correction_Adults-Children_UK

#Phenodata
phenodata <- read.delim(here('data','sample_information.txt'))

#Phenodata from batch1 (Older)
phenodata_batch1 <- read.delim(here('data', 'phenodata_batch1.tsv'))

#Effective gene size for FPKM normalization
gene_length <- read.delim(here('data', "GC_lengths.tsv"))
```

Normalize adjusted raw counts from Combat-seq with FPKM

```
rownames(genecounts_combat) <- genecounts_combat$Symbol

genecounts_combat <- as.data.frame(calculateFPKM(genecounts_combat[,-1],
                                gene_length[match(rownames(genecounts_combat),
                                                  rownames(gene_length)),1]))
```

Filter

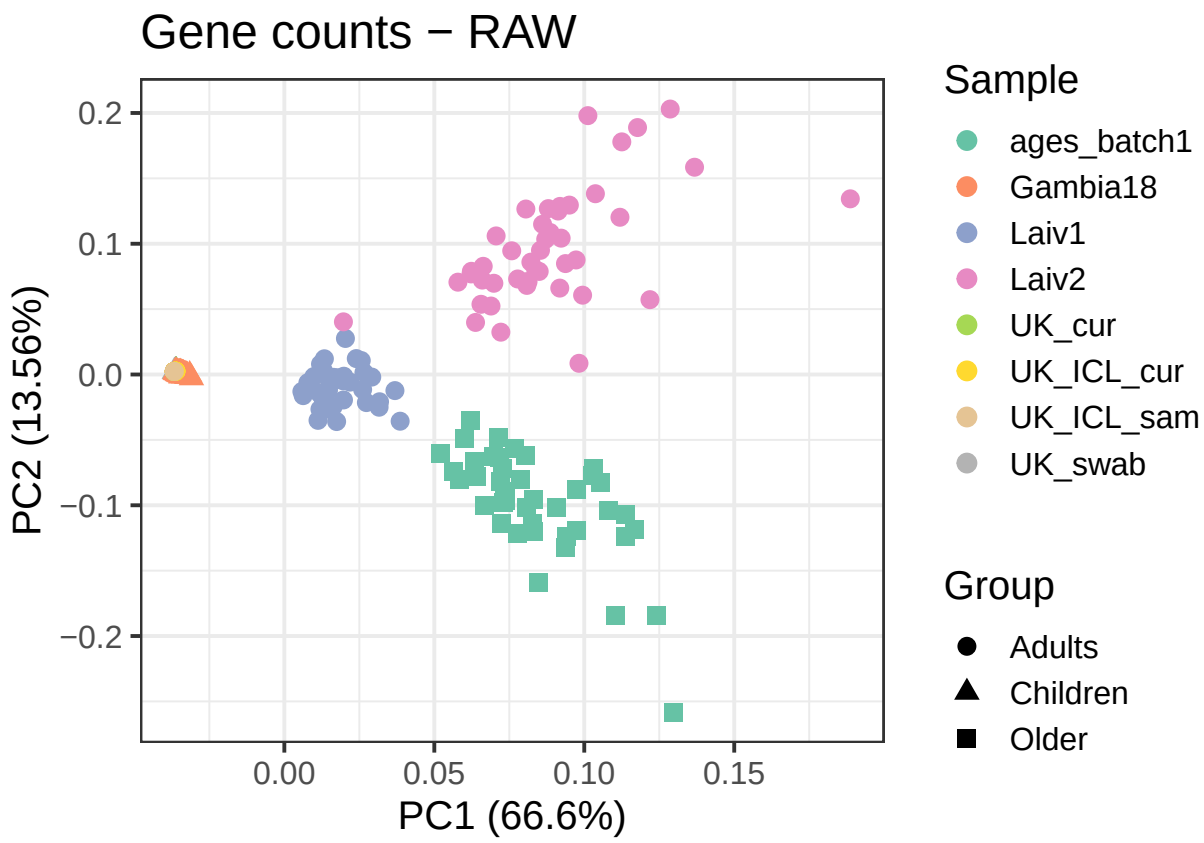
Use genes from "genecounts_fpkm" to all datasets. This genes show FPKM > 1 in more than 50% of samples

```
rownames(genecounts_raw) <- genecounts_raw$ID
gc_raw <- genecounts_raw[match(genecounts_fpkm$ID,
                               rownames(genecounts_raw)),-c(1:2)]

# remove samples D2 and D9 from batch1 (Older)
```

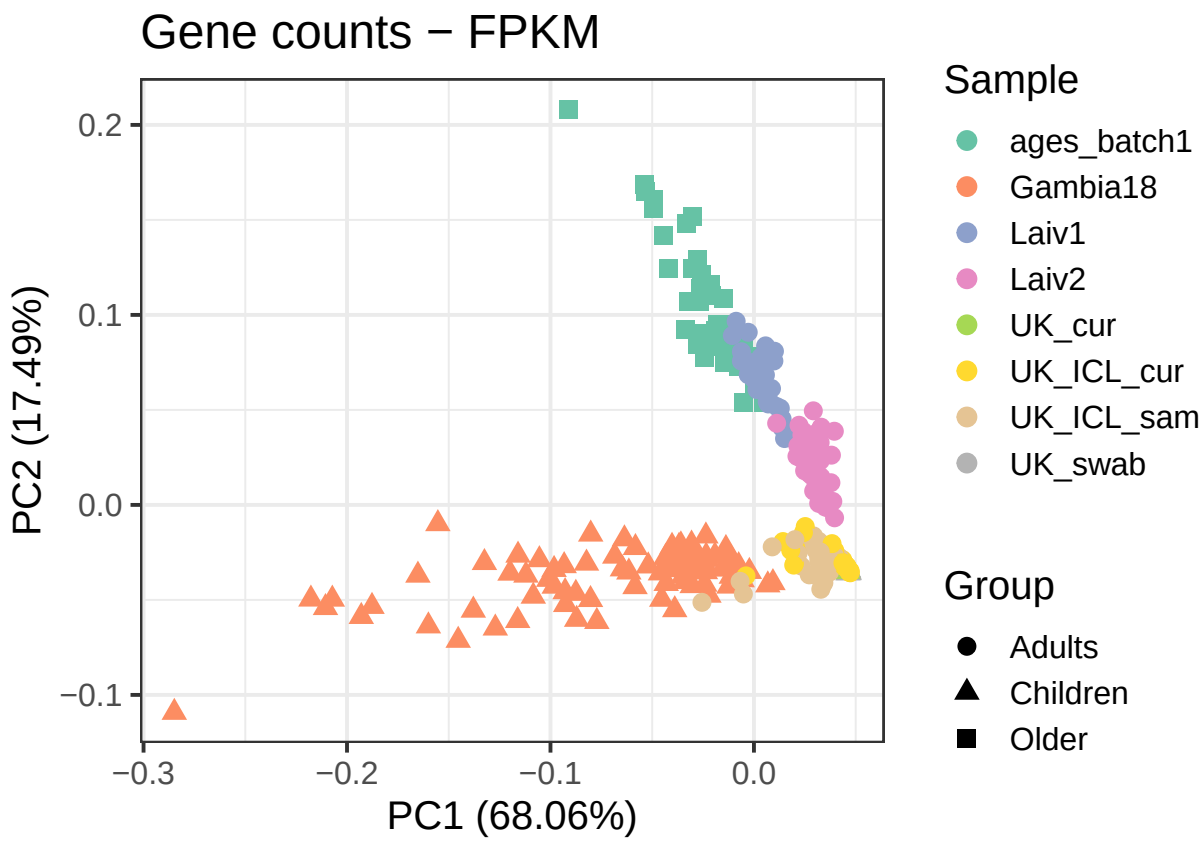
```r
gc_raw <- gc_raw[, !(colnames(gc_raw) %in% phenodata_batch1[phenodata_batch1$timepoint != 'baseline',1]]

gc_combat <- genecounts_combat[match(genecounts_fpkm$ID,
                                     rownames(genecounts_combat)),]

# remove samples D2 and D9 from batch1 (Older)
gc_combat <- gc_combat[, !(colnames(gc_combat) %in% phenodata_batch1[phenodata_batch1$timepoint != 'base

rownames(genecounts_scBatch) <- genecounts_scBatch$ID
gc_scbatch <- genecounts_scBatch[match(genecounts_fpkm$ID,
                                       rownames(genecounts_scBatch)),-c(1:2)]
# remove samples D2 and D9 from batch1 (Older)
gc_scbatch <- gc_scbatch[, !(colnames(gc_scbatch) %in% phenodata_batch1[phenodata_batch1$timepoint != '


rownames(genecounts_fpkm) <- genecounts_fpkm$ID
gc_fpkm <- genecounts_fpkm[,-c(1:2)]
```

PCA

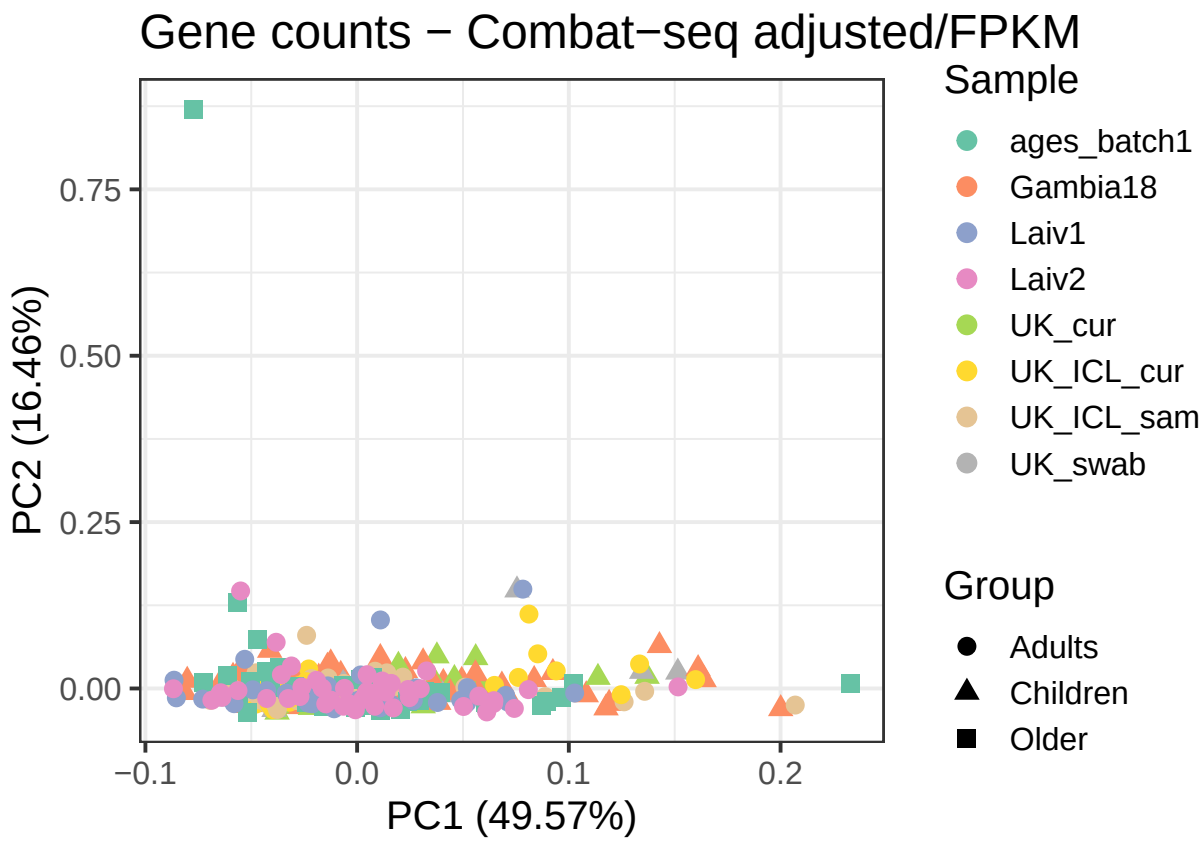
```r
### RAW
# Transpose and add Sample and Group from phenodata
t.gc_raw <-  as.data.frame(t(gc_raw))
t.gc_raw$label <- rownames(t.gc_raw)
t.gc_raw <- merge(phenodata, t.gc_raw, by = 'label')

#PCA
prcom.t.gc_raw <- prcomp(t.gc_raw[,-c(1:3)])

autoplot(prcom.t.gc_raw, t.gc_raw, shape = 'Group', colour= 'Sample', size=3) +
  scale_color_manual(values= brewer.pal(8,"Set2")) +
  theme_bw(base_size = 15)  +
  ggtitle("Gene counts - RAW")
```



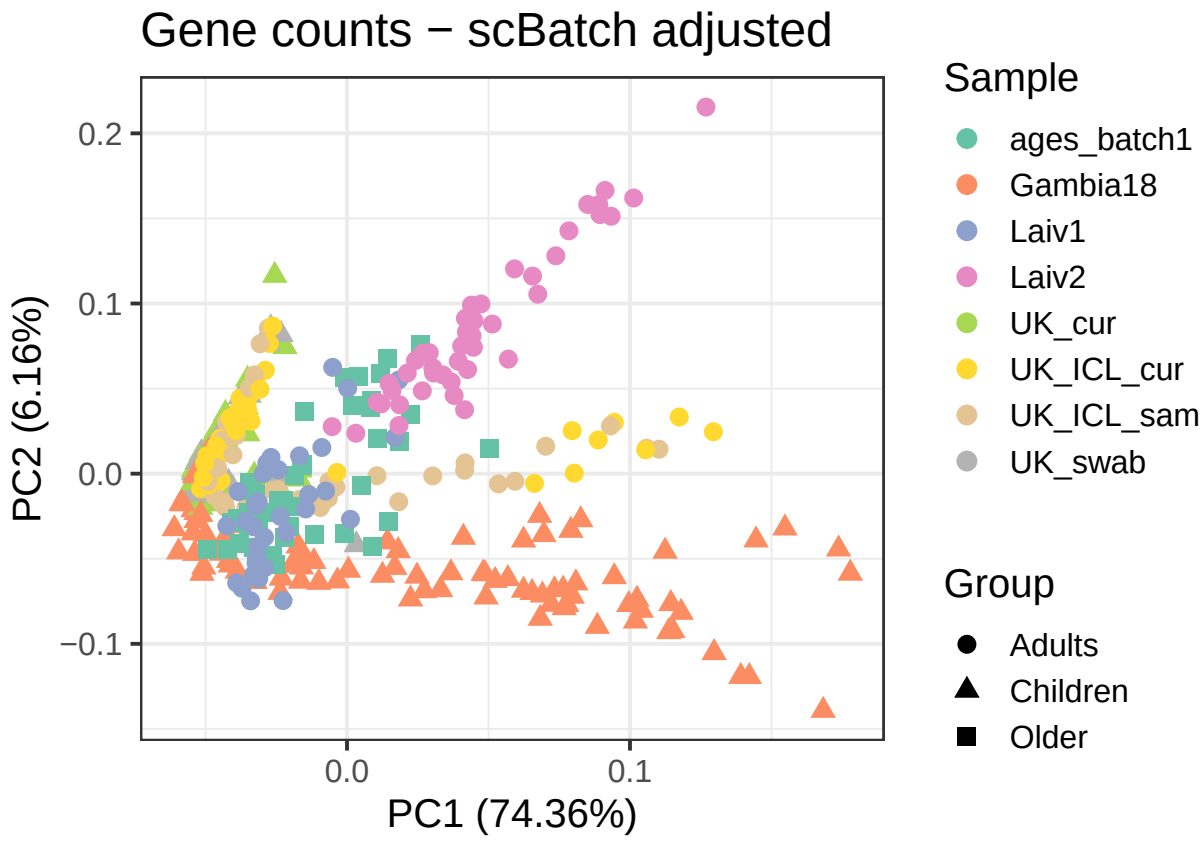

```
### FPKM
# Transpose and add Sample and Group from phenodata
t.gc_fpkm <-  as.data.frame(t(gc_fpkm))
t.gc_fpkm$label <- rownames(t.gc_fpkm)
t.gc_fpkm <- merge(phenodata, t.gc_fpkm, by = 'label')

#PCA
prcom.t.gc_fpkm <- prcomp(t.gc_fpkm[,-c(1:3)])

autoplot(prcom.t.gc_fpkm, t.gc_fpkm, shape = 'Group', colour= 'Sample', size=3) +
  scale_color_manual(values= brewer.pal(8,"Set2")) +
  theme_bw(base_size = 15) +
  ggtitle("Gene counts - FPKM")
```




```
### Combat-seq
# Transpose and add Sample and Group from phenodata
t.gc_combat <-  as.data.frame(t(gc_combat))
t.gc_combat$label <- rownames(t.gc_combat)
t.gc_combat <- merge(phenodata, t.gc_combat, by = 'label')

#PCA
prcom.t.gc_combat <- prcomp(t.gc_combat[,-c(1:3)])

autoplot(prcom.t.gc_combat, t.gc_combat, shape = 'Group', colour= 'Sample', size=3) +
  scale_color_manual(values= brewer.pal(8,"Set2")) +
  theme_bw(base_size = 15) +
  ggtitle("Gene counts - Combat-seq adjusted/FPKM")
```



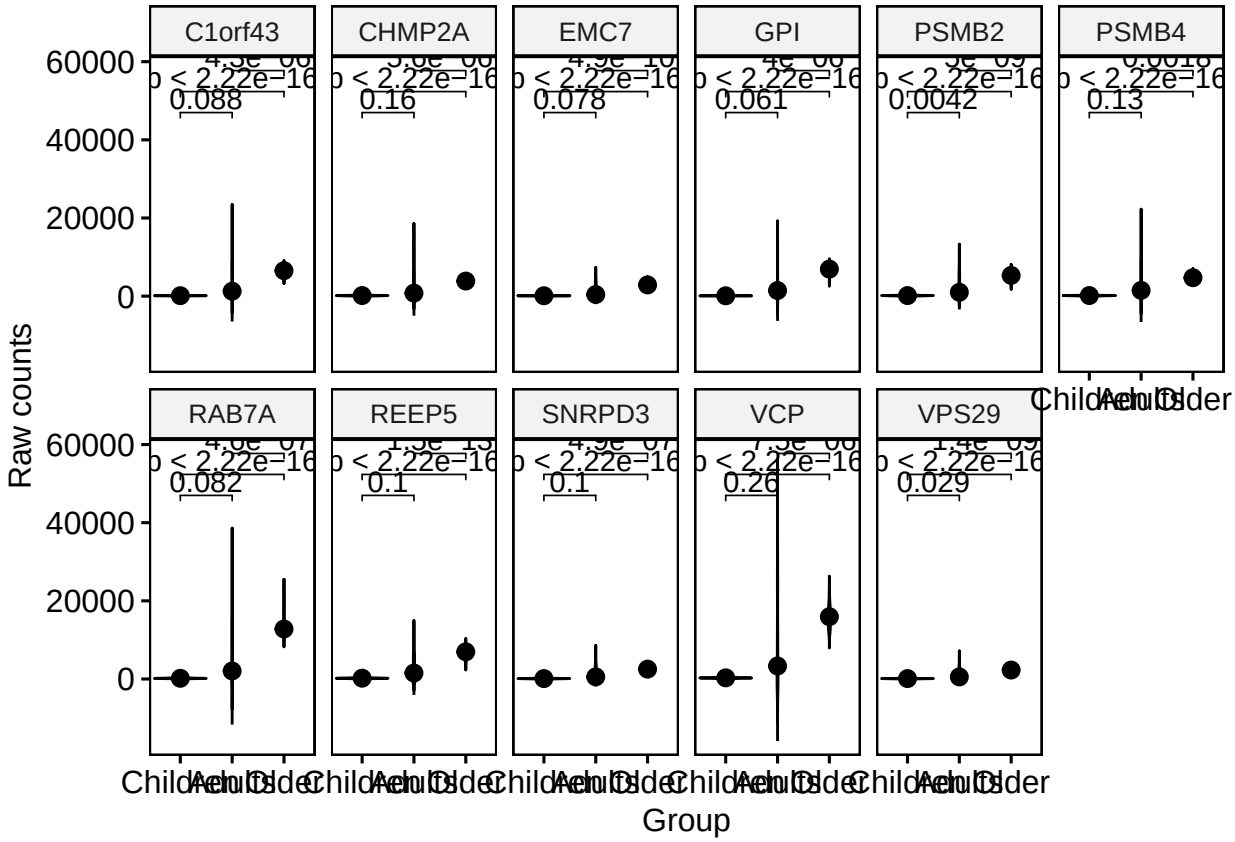

```
### scBatch
# Transpose and add Sample and Group from phenodata
t.gc_scbatch <-  as.data.frame(t(gc_scbatch))
t.gc_scbatch$label <- rownames(t.gc_scbatch)
t.gc_scbatch <- merge(phenodata, t.gc_scbatch, by = 'label')

#PCA
prcom.t.gc_scbatch <- prcomp(t.gc_scbatch[,-c(1:3)])

autoplot(prcom.t.gc_scbatch, t.gc_scbatch, shape = 'Group', colour= 'Sample', size=3) +
  scale_color_manual(values= brewer.pal(8,"Set2")) +
  theme_bw(base_size = 15) +
  ggtitle("Gene counts - scBatch adjusted")
```



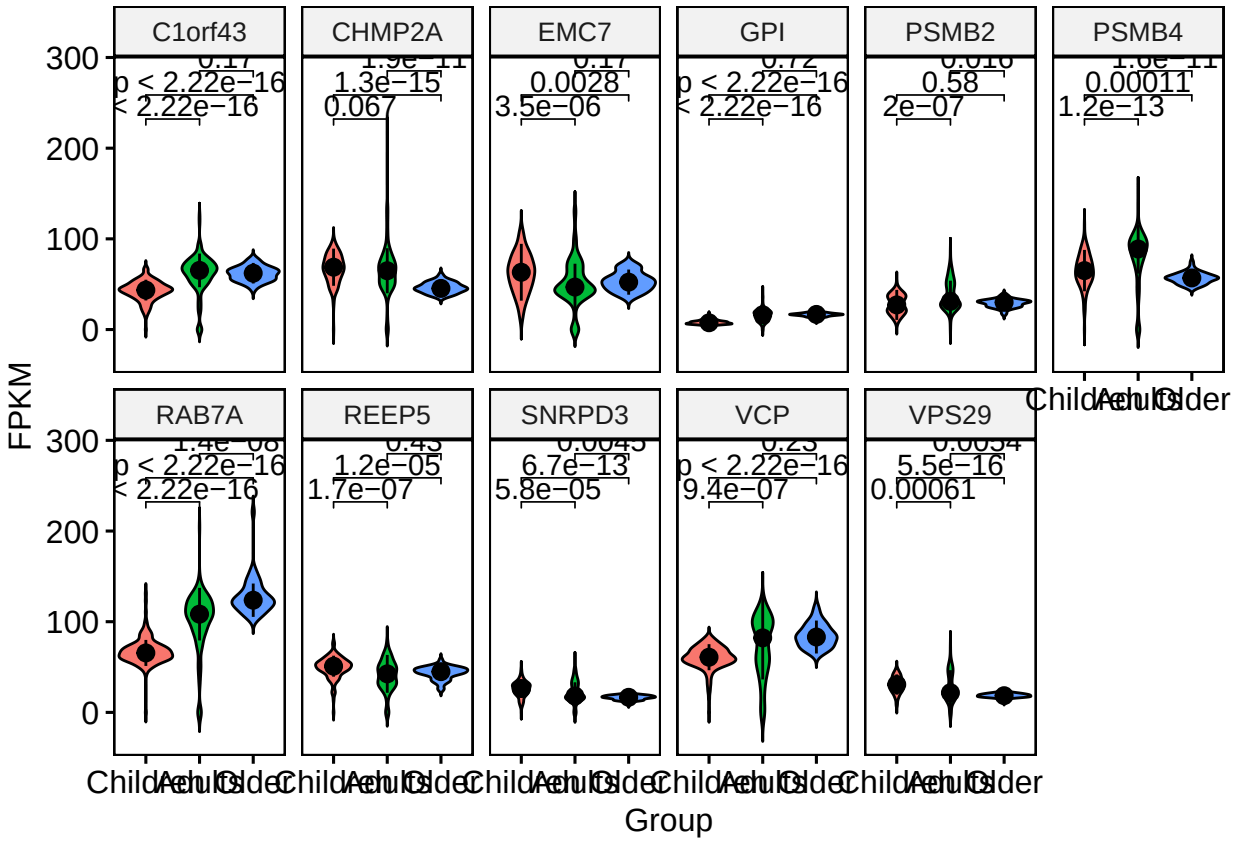
Plot housekeeping genes expression

```
#House keeping genes dataframe
hkg_genesymbol <- c("C1orf43","CHMP2A","EMC7","GPI",
                    "PSMB2","PSMB4","RAB7A","REEP5",
                    "SNRPD3","VCP","VPS29")

hkg_ensemblid <- c("ENSG00000143612", "ENSG00000130724", "ENSG00000134153", "ENSG00000105220",
                   "ENSG00000126067", "ENSG00000159377", "ENSG00000075785", "ENSG00000129625",
                   "ENSG00000100028", "ENSG00000165280", "ENSG00000111237")

hkg_df <- data.frame(genesymbol = hkg_genesymbol, ensemblid = hkg_ensemblid)


hkg_raw <- cbind(t.gc_raw[,1:3] , t.gc_raw[ , colnames(t.gc_raw) %in% hkg_df$ensemblid])
hkg_raw <- tidyr::gather(hkg_raw, "Genes", "Counts", 4:14)
hkg_raw$Group <- relevel(hkg_raw$Group, 'Children')
hkg_raw <- merge(hkg_raw, hkg_df, by.x='Genes', by.y='ensemblid')

hkg_fpkm <- cbind(t.gc_fpkm[,1:3] , t.gc_fpkm[ , colnames(t.gc_fpkm) %in% hkg_df$ensemblid])
hkg_fpkm <- tidyr::gather(hkg_fpkm, "Genes", "Counts", 4:14)
hkg_fpkm$Group <- relevel(hkg_fpkm$Group, 'Children')
hkg_fpkm <- merge(hkg_fpkm, hkg_df, by.x='Genes', by.y='ensemblid')

hkg_combat <- cbind(t.gc_combat[,1:3] , t.gc_combat[ , colnames(t.gc_combat) %in% hkg_df$ensemblid])
```

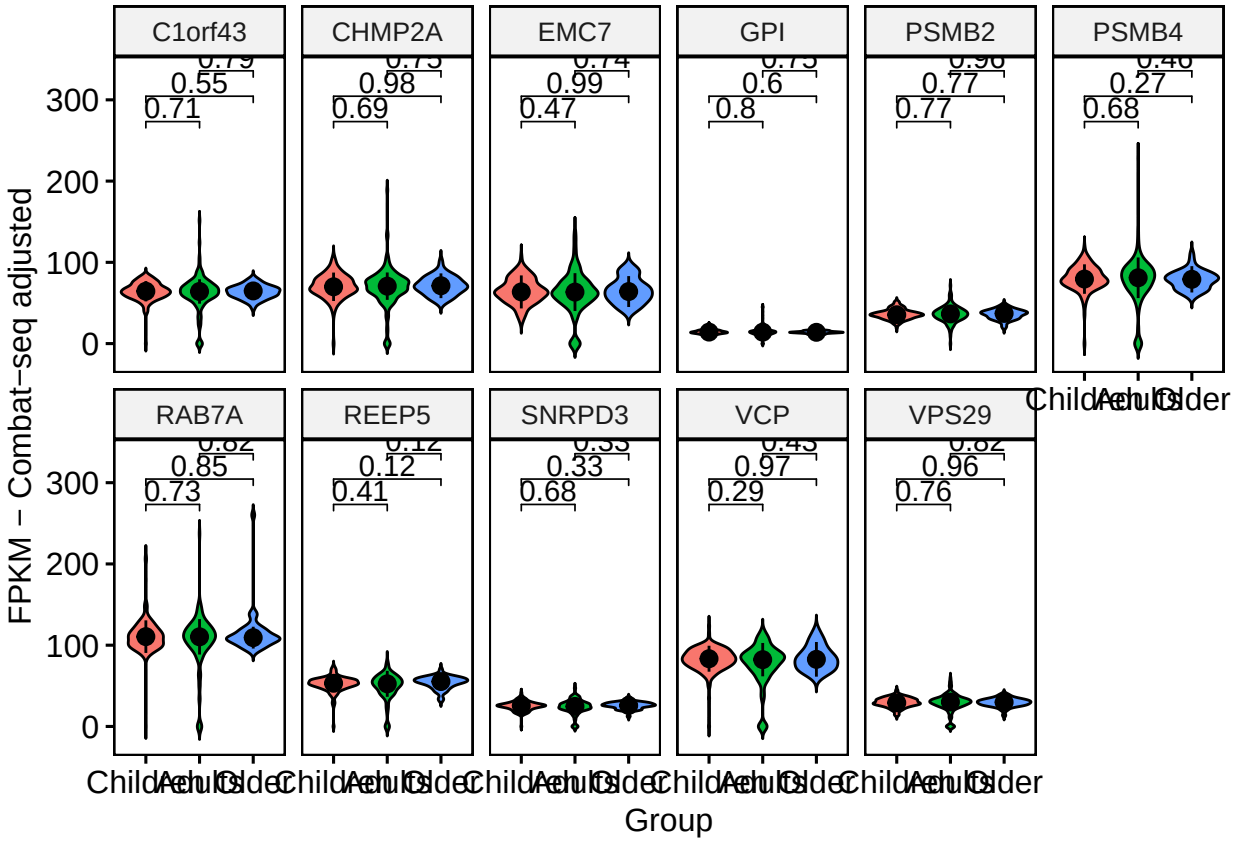
```
hkg_combat <- tidyr::gather(hkg_combat, "Genes", "Counts", 4:14)
hkg_combat$Group <- relevel(hkg_combat$Group, 'Children')
hkg_combat <- merge(hkg_combat, hkg_df, by.x='Genes', by.y='ensemblid')

hkg_scbatch <- cbind(t.gc_scbatch[,1:3] , t.gc_scbatch[ , colnames(t.gc_scbatch) %in% hkg_df$ensemblid])
hkg_scbatch <- tidyr::gather(hkg_scbatch, "Genes", "Counts", 4:14)
hkg_scbatch$Group <- relevel(hkg_scbatch$Group, 'Children')
hkg_scbatch <- merge(hkg_scbatch, hkg_df, by.x='Genes', by.y='ensemblid')


my_comparisons <- list( c("Children", "Adults"), c("Children", "Older"),
                        c("Adults", "Older") )

## Violin-plot Raw
ggviolin(hkg_raw, x='Group',y = 'Counts',
         facet.by = 'genesymbol',
         color='black',
         legend= '',
         nrow=2,
         add='median_iqr',
         fill='Group',
         add.params = list(fill = "white"),
         ylab = 'Raw counts') +
 stat_compare_means(comparisons = my_comparisons)
```



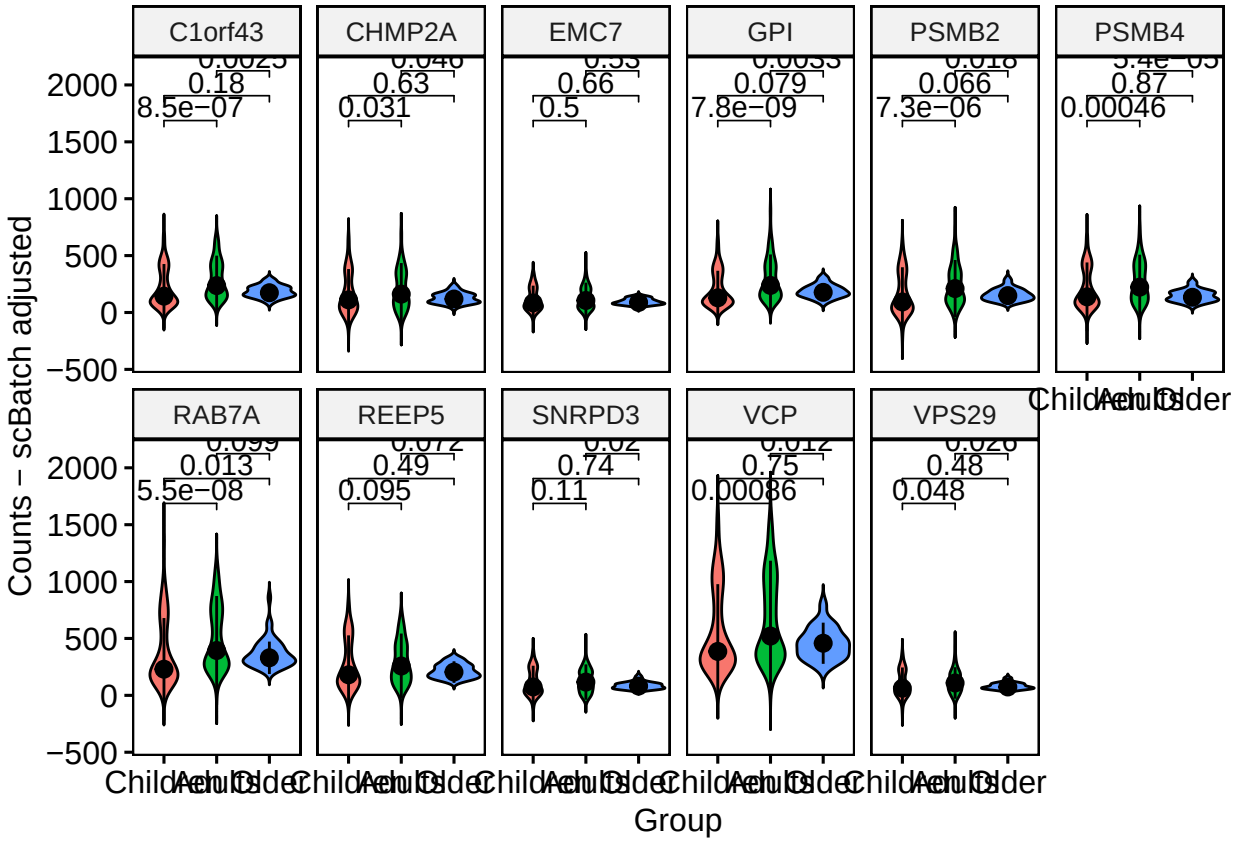
```
# scale_y_continuous(limits = c(1, 11))

## Violin-plot FPKM
ggviolin(hkg_fpkm, x='Group',y = 'Counts',
         facet.by = 'genesymbol',
         color='black',
         legend= '',
         nrow=2,
         add='median_iqr',
         fill='Group',
         add.params = list(fill = "white"),
         ylab = 'FPKM') +
 stat_compare_means(comparisons = my_comparisons)
```




```
## Violin-plot Combat
ggviolin(hkg_combat, x='Group',y = 'Counts',
         facet.by = 'genesymbol',
         color='black',
         legend= '',
         nrow=2,
         add='median_iqr',
         fill='Group',
         add.params = list(fill = "white"),
         ylab = 'FPKM - Combat-seq adjusted') +
 stat_compare_means(comparisons = my_comparisons)
```



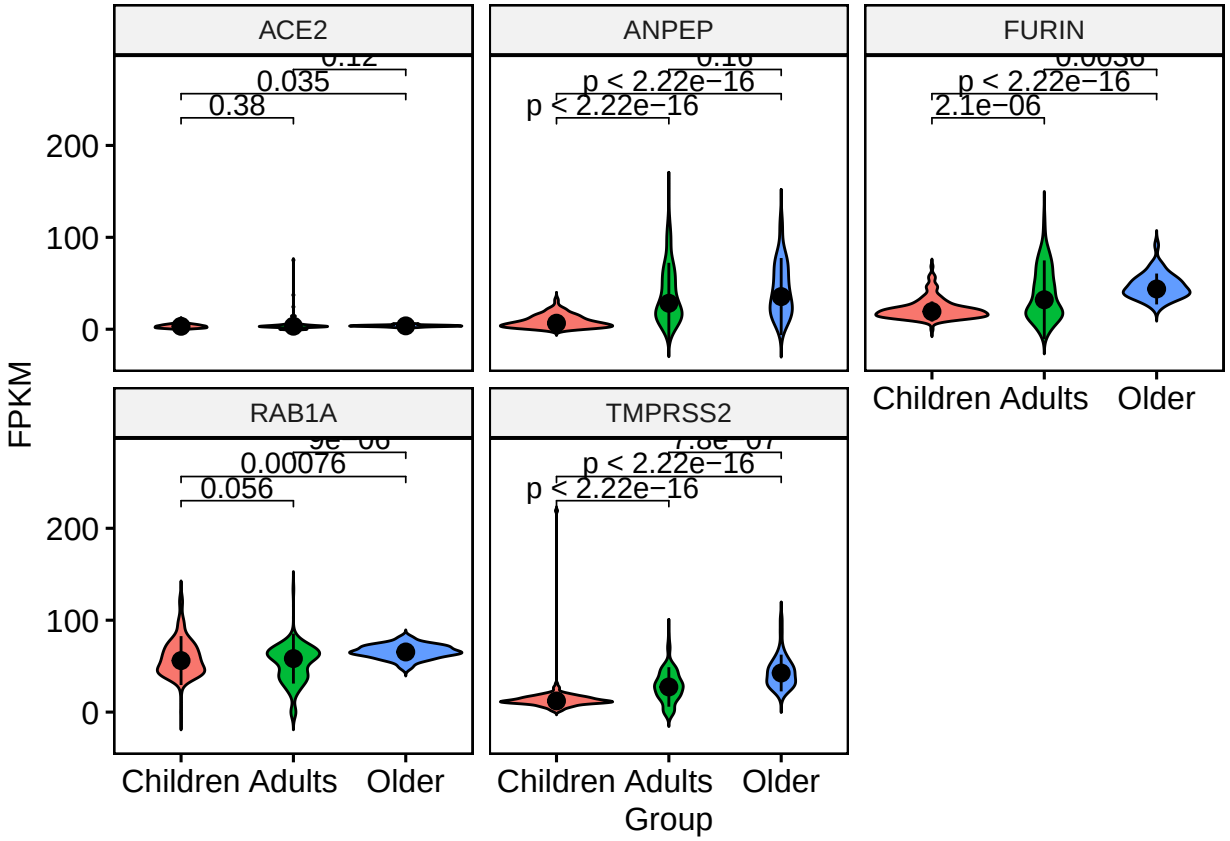

```
## Violin-plot scBatch
ggviolin(hkg_scbatch, x='Group',y = 'Counts',
         facet.by = 'genesymbol',
         color='black',
         legend= '',
         nrow=2,
         add='median_iqr',
         fill='Group',
         add.params = list(fill = "white"),
         ylab = 'Counts - scBatch adjusted') +
 stat_compare_means(comparisons = my_comparisons)
```



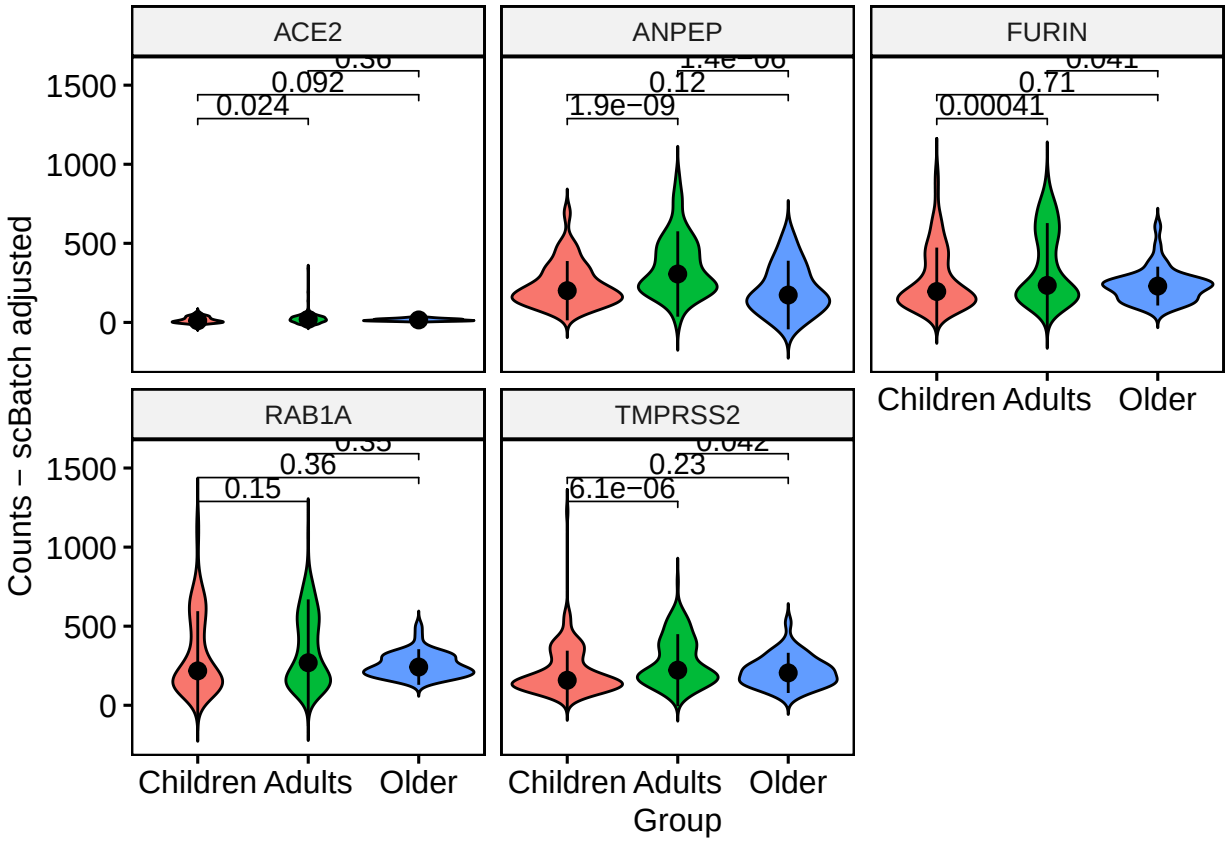
Expression analysis selected genes

```
#Selected genes dataframe
sg_genesymbol <- c("ACE2","TMPRSS2","FURIN","DPP4",
                   "ANPEP","RAB1A")

sg_ensemblid <- c("ENSG00000130234", "ENSG00000184012", "ENSG00000140564", "ENSG00000197635",
                  "ENSG00000166825", "ENSG00000138069")

sg_df <- data.frame(genesymbol = sg_genesymbol, ensemblid = sg_ensemblid)

sg_fpkm <- cbind(t.gc_fpkm[,1:3] , t.gc_fpkm[ , colnames(t.gc_fpkm) %in% sg_df$ensemblid])
sg_fpkm <- tidyr::gather(sg_fpkm, "Genes", "Counts", 4:8)
sg_fpkm$Group <- relevel(sg_fpkm$Group, 'Children')
sg_fpkm <- merge(sg_fpkm, sg_df, by.x='Genes', by.y='ensemblid')
#sg_fpkm$Counts <- log2(sg_fpkm$Counts)

sg_scbatch <- cbind(t.gc_scbatch[,1:3] , t.gc_scbatch[ , colnames(t.gc_scbatch) %in% sg_df$ensemblid])
sg_scbatch <- tidyr::gather(sg_scbatch, "Genes", "Counts", 4:8)
sg_scbatch$Group <- relevel(sg_scbatch$Group, 'Children')
sg_scbatch <- merge(sg_scbatch, sg_df, by.x='Genes', by.y='ensemblid')

## Violin-plot FPKM
ggviolin(sg_fpkm, x='Group',y = 'Counts',
         facet.by = 'genesymbol',
```

```
         color='black',
         legend= '',
         nrow=2,
         add='median_iqr',
         fill='Group',
         add.params = list(fill = "white"),
         ylab = 'FPKM') +
 stat_compare_means(comparisons = my_comparisons)
```




```
# scale_y_continuous(limits = c(-2, 13))


## Violin-plot scBatch
ggviolin(sg_scbatch, x='Group',y = 'Counts',
         facet.by = 'genesymbol',
         color='black',
         legend= '',
         nrow=2,
         add="median_iqr",
         fill='Group',
         add.params = list(fill = "white"),
         ylab = 'Counts - scBatch adjusted') +
 stat_compare_means(comparisons = my_comparisons)
```


ACE2
ANPEP
FURIN
RAB1A
TMPRSS2
Counts – scBatch adjusted
1500
1000
500
0
0.024
0.092
0.12
1.9e−09
0.00041
0.71
0.15
0.36
0.23
6.1e−06
Children
Adults
Older
Group