Cluster statistics

```
library(Seurat)
library(ggplot2)
library(reshape2)
library(plyr)
library(RColorBrewer)
library(pheatmap)


subdataset <- readRDS('../tmp/crc_smc.malignantcells.Rds')
head(subdataset@meta.data, n=3); nrow(subdataset@meta.data)
```

```
##                                   nCount_RNA nFeature_RNA        Library
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1      35998         4823 PM-PS-0001-T-A1
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1      31383         5252 PM-PS-0001-T-A1
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1       7302         1713 PM-PS-0001-T-A1
##                                   Patient  Sample Cell_subtype RNA_snn_res.0.3
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1  SMC01 SMC01-T         CMS2               2
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1  SMC01 SMC01-T         CMS2               0
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1  SMC01 SMC01-T         CMS2               2
```

```
## [1] 17334
```

```
### Proportion of patients in each cluster
prop <- data.frame(matrix(ncol = length(levels(subdataset@meta.data$RNA_snn_res.0.3)), nrow = length(le
colnames(prop) <- as.character(levels(subdataset@meta.data$RNA_snn_res.0.3))
rownames(prop) <- levels(subdataset@meta.data$Patient)

for (cluster in levels(subdataset@meta.data$RNA_snn_res.0.3)) {
  tmp <- subset(subdataset@meta.data, RNA_snn_res.0.3 == cluster)
  tmp_prop <- summary(tmp$Patient)/sum(summary(tmp$Patient)) *100
  prop[, cluster] <- tmp_prop
}

prop_t <- melt(data.frame(Patient = rownames(prop), prop, check.rows = F, check.names = F))
```

```
## Using Patient as id variables
```

```
head(prop_t)
```

```
##   Patient variable    value
## 1   SMC01        0 5.976698
## 2   SMC02        0 2.663035
## 3   SMC03        0 5.053715
## 4   SMC04        0 1.482826
## 5   SMC06        0 1.407172
## 6   SMC07        0 2.330156
```

```
getPalette <- colorRampPalette(brewer.pal(11, 'Spectral'))

ggplot(prop_t, aes(variable, value, fill = Patient)) +
  geom_bar(stat = 'identity', position = 'stack') +
  scale_fill_manual(values = getPalette(22)) +
  labs(x = 'Cluster', y = 'Proportion (%)') +
  theme_bw() +
  theme(panel.grid = element_blank(),
        axis.text = element_text(colour = 'black'),
        axis.ticks = element_line(colour = 'black'))
```
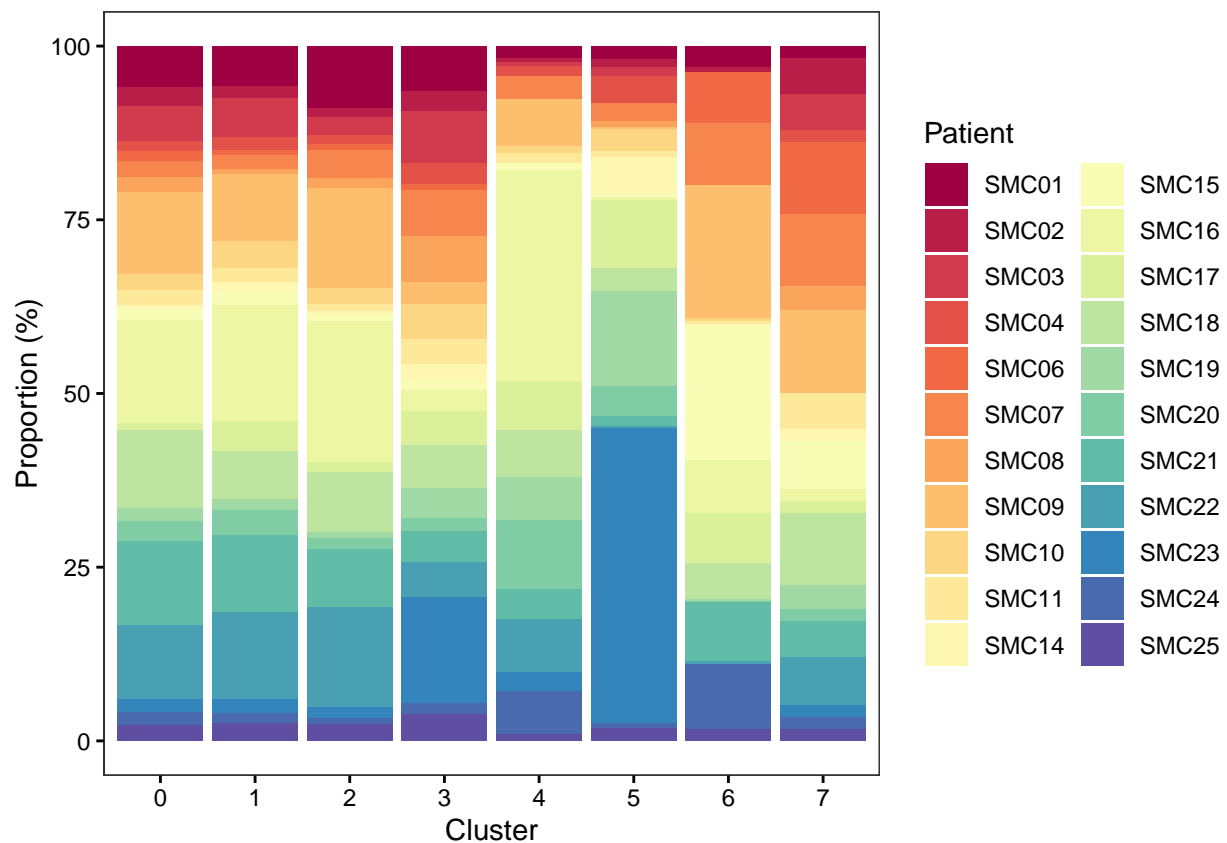


```
#ggsave('../res_various/res_0.3/prop_res_0.3.pdf', units = 'cm', width = 12, height = 9)

### The number of cells in each cluster
cellcount <- data.frame(Size = summary(subdataset@meta.data$RNA_snn_res.0.3))
cellcount <- data.frame(res = rownames(cellcount), cellcount, check.rows = F)
cellcount$res <- factor(cellcount$res, levels = levels(subdataset@meta.data$RNA_snn_res.0.3))

ggplot(cellcount, aes(res, Size)) +
  geom_bar(stat = 'identity', position = 'stack') +
  labs(x = 'Cluster', y = 'Cell count') +
  theme_bw() +
  theme(panel.grid = element_blank(),
        axis.text = element_text(colour = 'black'),
        axis.ticks = element_line(colour = 'black'))
```
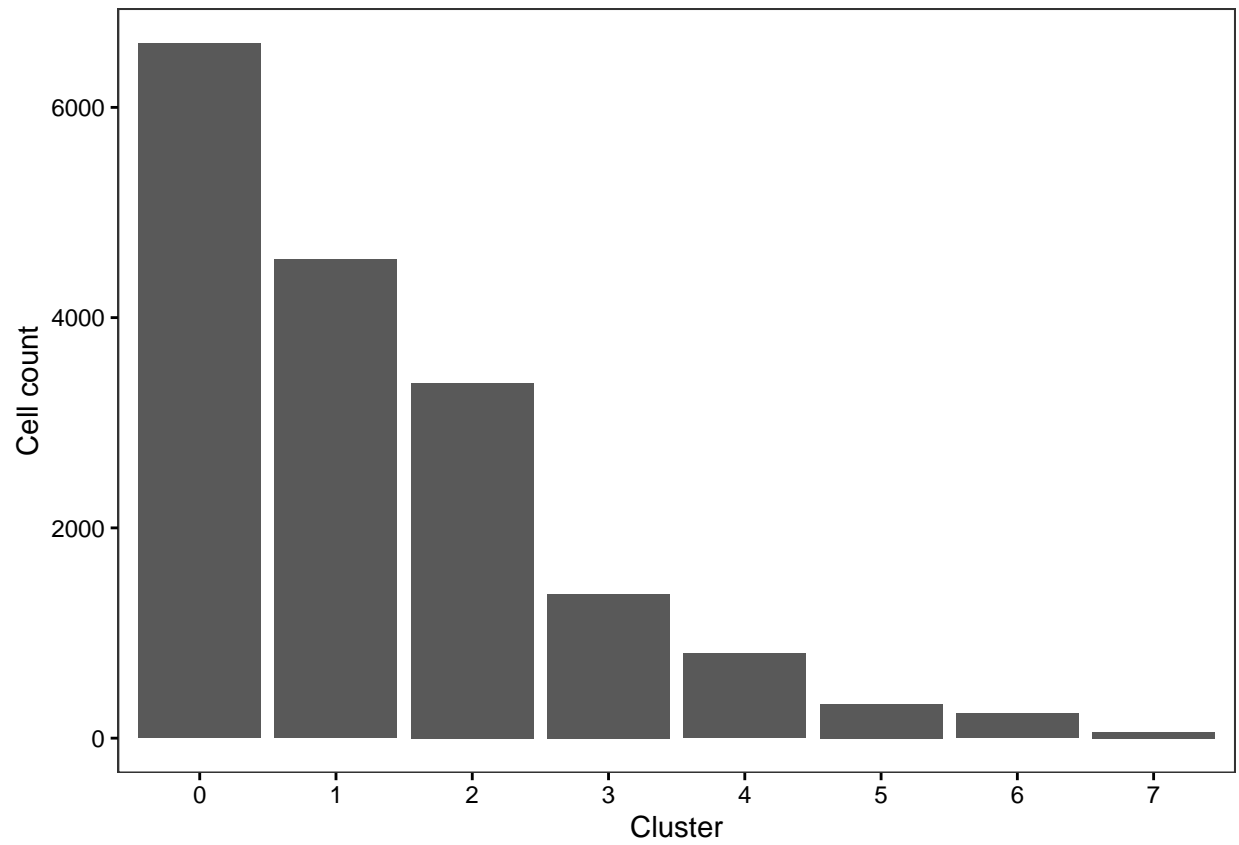
```
#ggsave('../res_various/res_0.3/count_res_0.3.pdf', units = 'cm', width = 8, height = 9)
```