

Correlation b/w malignant cluster and cell types

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

#dir.create('../res_various/res_0.3/cluster_proportion')

myObj <- readRDS('../tmp/crc_smc.malignantcells.Rds')
myObj@meta.data$RNA_snn_res.0.3 <- paste('C', myObj@meta.data$RNA_snn_res.0.3, sep = '')
myObj@meta.data$RNA_snn_res.0.3 <- factor(myObj@meta.data$RNA_snn_res.0.3, levels = paste('C', c(0:7),
levels(myObj@meta.data$RNA_snn_res.0.3)

## [1] "C0" "C1" "C2" "C3" "C4" "C5" "C6" "C7"

head(myObj@meta.data, n=3)

##                               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             C2
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1  SMC01  SMC01-T      CMS2             C0
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1  SMC01  SMC01-T      CMS2             C2

nrow(myObj@meta.data) # 17334

## [1] 17334

malignant <- myObj@meta.data
remove(myObj)

### Full data
fullObj <- readRDS('../tmp/crc_smc.full.Rds')
excludecells <- setdiff(rownames(subset(fullObj@meta.data, Cell_type == 'Malignant cells')), rownames(m

## [1] 135

fullObj <- subset(fullObj, cells = setdiff(rownames(fullObj@meta.data), excludecells))
fullObj@meta.data$Cell_type_v2 <- mapvalues(fullObj@meta.data$Cell_subtype, from = levels(fullObj@meta.
to = c("B cells", "Plasma cells", "Plasma cells",
```

```

"Epithelial cells", "Epithelial cells", "Epithelial cells", "Epithelial cells", "Epithelial cells",
"Macrophages", "Macrophages", "Macrophages", "Macrophages", "Macrophages", "Macrophages",
"Fibroblasts", "Fibroblasts", "Fibroblasts", "Fibroblasts", "Fibroblasts", "Fibroblasts",
"Stromal cells", "Stromal cells",
"Endothelial cells", "Endothelial cells", "Endothelial cells", "Endothelial cells", "Endothelial cells",
"Malignant cells", "Malignant cells", "Malignant cells", "Malignant cells", "Malignant cells",
"Epithelial cells", "Epithelial cells", "Epithelial cells", "Epithelial cells", "Epithelial cells",
fullObj@meta.data$Cell_type_v2 <- factor(fullObj@meta.data$Cell_type_v2,
levels = c("Epithelial cells", "Malignant cells", "Macrophages",
"B cells", "Plasma cells", "T cells", "Fibroblasts",
summary(fullObj@meta.data$Cell_type_v2)

```

```

## Epithelial cells Malignant cells Macrophages cDC
## 1070 17334 5822 482
## Mast cells B cells Plasma cells T cells
## 187 3733 5301 21511
## Fibroblasts Stromal cells Endothelial cells
## 3462 964 1507

```

```
head(fullObj@meta.data, n=3)
```

```

## orig.ident nCount_RNA nFeature_RNA
## AAACCTGGTGTAATGA-1-PM-PS-0001-N-A1 ColonCancer 2911 969
## AAACCTGTACCTCGT-1-PM-PS-0001-N-A1 ColonCancer 5531 1481
## AAACGGGCAACACCTA-1-PM-PS-0001-N-A1 ColonCancer 2565 501
## Library Patient Class Sample
## AAACCTGGTGTAATGA-1-PM-PS-0001-N-A1 PM-PS-0001-N-A1 SMC01 Normal SMC01-N
## AAACCTGTACCTCGT-1-PM-PS-0001-N-A1 PM-PS-0001-N-A1 SMC01 Normal SMC01-N
## AAACGGGCAACACCTA-1-PM-PS-0001-N-A1 PM-PS-0001-N-A1 SMC01 Normal SMC01-N
## Cell_type Cell_subtype percent.mt
## AAACCTGGTGTAATGA-1-PM-PS-0001-N-A1 T cells CD4+ T cells 5.565098
## AAACCTGTACCTCGT-1-PM-PS-0001-N-A1 T cells CD4+ T cells 2.332309
## AAACGGGCAACACCTA-1-PM-PS-0001-N-A1 B cells IgA+ Plasma 16.179337
## Cell_type_v2
## AAACCTGGTGTAATGA-1-PM-PS-0001-N-A1 T cells
## AAACCTGTACCTCGT-1-PM-PS-0001-N-A1 T cells
## AAACGGGCAACACCTA-1-PM-PS-0001-N-A1 Plasma cells

```

```

full_label <- fullObj@meta.data
remove(fullObj)

```

```
### Proportions of malignant clusters
```

```

malignant <- subset(malignant, `RNA_snn_res.0.3` != 'C7')
malignant <- droplevels(malignant)
levels(malignant$RNA_snn_res.0.3)

```

```
## [1] "C0" "C1" "C2" "C3" "C4" "C5" "C6"
```

```

prop <- data.frame(matrix(nrow = length(unique(malignant$RNA_snn_res.0.3)), ncol = length(unique(malignant$Patient)),
colnames(prop) <- levels(malignant$Patient)
rownames(prop) <- levels(malignant$RNA_snn_res.0.3)
prop

```

```
##      SMC01 SMC02 SMC03 SMC04 SMC06 SMC07 SMC08 SMC09 SMC10 SMC11 SMC14 SMC15
## C0      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C1      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C2      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C3      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C4      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C5      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C6      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
##      SMC16 SMC17 SMC18 SMC19 SMC20 SMC21 SMC22 SMC23 SMC24 SMC25
## C0      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C1      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C2      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C3      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C4      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C5      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## C6      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
```

```
for (pt in levels(malignant$Patient)) {
  summ <- summary(subset(malignant, Patient == pt)$RNA_snn_res.0.3)
  prop[, pt] <- round(summ/sum(summ)*100, digits = 2)
}
prop
```

```
##      SMC01 SMC02 SMC03 SMC04 SMC06 SMC07 SMC08 SMC09 SMC10 SMC11 SMC14 SMC15
## C0 36.74 50.87 42.49 32.24 51.67 29.39 43.53 42.28 31.08 42.37 18.40 38.26
## C1 24.28 22.25 32.44 29.93 17.22 17.18 11.36 23.70 35.26 27.73 41.10 26.37
## C2 28.28 11.85 11.07 15.46 15.00 25.95 15.14 26.24 15.74 10.28 9.82 10.93
## C3 8.19 11.85 12.98 14.14 6.67 16.98 28.71 2.38 13.94 14.95 20.25 6.11
## C4 1.30 1.45 0.51 3.95 0.00 4.96 0.32 2.92 1.79 3.43 0.00 2.89
## C5 0.56 1.16 0.51 4.28 0.00 1.53 0.95 0.05 1.99 0.93 10.43 0.64
## C6 0.65 0.58 0.00 0.00 9.44 4.01 0.00 2.43 0.20 0.31 0.00 14.79
##      SMC16 SMC17 SMC18 SMC19 SMC20 SMC21 SMC22 SMC23 SMC24 SMC25
## C0 35.74 14.4 49.20 33.16 35.00 47.03 37.31 19.47 37.26 36.50
## C1 27.90 40.6 20.62 19.36 31.73 29.31 29.97 14.60 21.66 27.49
## C2 25.17 10.2 19.36 6.90 10.19 16.61 25.70 7.85 10.83 18.98
## C3 1.51 13.6 5.64 15.38 5.00 3.55 3.64 33.12 6.37 12.90
## C4 8.98 11.4 3.65 13.26 15.38 2.04 3.27 3.30 16.24 1.70
## C5 0.04 6.4 0.73 11.67 2.69 0.29 0.05 21.66 0.64 1.46
## C6 0.66 3.4 0.80 0.27 0.00 1.17 0.05 0.00 7.01 0.97
```

```
prop_plot <- data.frame(Cluster = rownames(prop), prop)
prop_plot <- melt(prop_plot)
```

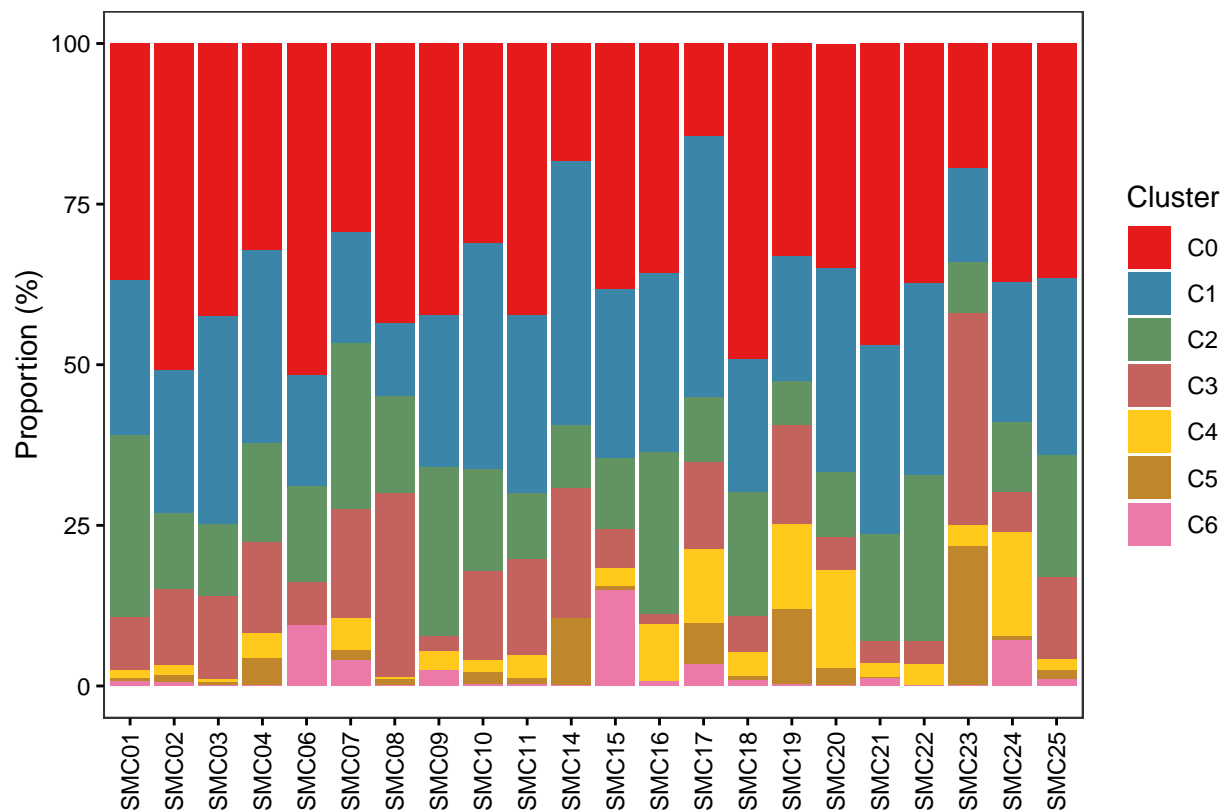
```
## Using Cluster as id variables
```

```
prop_plot$Cluster <- factor(prop_plot$Cluster, levels = levels(malignant$RNA_snn_res.0.3))
head(prop_plot)
```

```
##      Cluster variable value
## 1      C0      SMC01 36.74
## 2      C1      SMC01 24.28
## 3      C2      SMC01 28.28
```

```
## 4      C3      SMC01  8.19
## 5      C4      SMC01  1.30
## 6      C5      SMC01  0.56
```

```
getPalette <- colorRampPalette(brewer.pal(9, 'Set1'))
ggplot(prop_plot, aes(variable, value, fill = Cluster)) +
  geom_bar(stat = 'identity', position = 'stack') +
  scale_fill_manual(values = getPalette(8)[-8]) +
  labs(x = '', y = 'Proportion (%)') +
  theme_bw() +
  theme(panel.grid = element_blank(),
        axis.text = element_text(color = 'black'),
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5),
        axis.ticks = element_line(color = 'black'))
```



```
#ggsave('../res_various/res_0.3/cluster_proportion/prop.clustering.pdf', unit = 'cm', width = 12, height = 12)
```

```
### Proportions of cell types
```

```
full_label <- subset(full_label, Class == 'Tumor')
full_label <- subset(full_label, Patient != 'SMC05') # minor library
summary(full_label$Cell_type_v2)
```

```
## Epithelial cells      Malignant cells      Macrophages      cDC
##                   0          17334          5571          352
```

```
##      Mast cells      B cells      Plasma cells      T cells
##           3          1972          1813          14741
##      Fibroblasts      Stromal cells Endothelial cells
##           1493          462           765
```

```
full_label <- subset(full_label, Cell_type_v2 != 'Mast cells') # only three cells in tumor samples
summary(full_label$Cell_type_v2)
```

```
## Epithelial cells      Malignant cells      Macrophages      cDC
##           0          17334          5571          352
##      Mast cells      B cells      Plasma cells      T cells
##           0          1972          1813          14741
##      Fibroblasts      Stromal cells Endothelial cells
##           1493          462           765
```

```
full_label <- droplevels(full_label)
```

```
prop_ct <- data.frame(matrix(nrow = length(unique(full_label$Cell_type_v2)), ncol = length(unique(full_label$Patient)),
  colnames(prop_ct) <- levels(full_label$Patient)
  rownames(prop_ct) <- levels(full_label$Cell_type_v2)
  head(prop_ct)
```

```
##      SMC01 SMC02 SMC03 SMC04 SMC06 SMC07 SMC08 SMC09 SMC10 SMC11
## Malignant cells      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## Macrophages          NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## cDC                  NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## B cells              NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## Plasma cells         NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## T cells              NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
##      SMC14 SMC15 SMC16 SMC17 SMC18 SMC19 SMC20 SMC21 SMC22 SMC23
## Malignant cells      NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## Macrophages          NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## cDC                  NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## B cells              NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## Plasma cells         NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
## T cells              NA      NA      NA      NA      NA      NA      NA      NA      NA      NA
##      SMC24 SMC25
## Malignant cells      NA      NA
## Macrophages          NA      NA
## cDC                  NA      NA
## B cells              NA      NA
## Plasma cells         NA      NA
## T cells              NA      NA
```

```
for (pt in levels(full_label$Patient)) {
  summ <- summary(subset(full_label, Patient == pt)$Cell_type_v2)
  prop_ct[, pt] <- round(summ/sum(summ)*100, digits = 2)
}
prop_ct
```

```
##      SMC01 SMC02 SMC03 SMC04 SMC06 SMC07 SMC08 SMC09 SMC10 SMC11
## Malignant cells      41.80 22.02 39.53 16.98 16.62 13.33 17.14 49.89 23.87 19.27
```

```
## Macrophages      5.01 34.57 15.08 42.20 18.23  6.14 14.88  9.45 27.01  3.57
## cDC              0.35  1.96  0.35  1.67  0.45  0.80  0.64  0.43  0.90  0.48
## B cells          2.49  4.42  7.97  5.07  2.68  7.09  8.76  0.24  1.05  3.63
## Plasma cells     2.45  1.14  6.96  0.61  6.26 13.10  6.93  0.32  4.28 11.24
## T cells          41.92 29.27 19.19 25.17 48.35 54.29 43.15 38.46 35.38 58.66
## Fibroblasts      3.61  4.04  9.02  2.17  5.99  1.66  6.77  0.38  4.42  1.31
## Stromal cells    0.74  0.38  0.30  1.61  0.98  1.18  0.43  0.30  1.19  0.59
## Endothelial cells 1.63  2.21  1.60  4.51  0.45  2.41  1.29  0.54  1.90  1.25
## SMC14 SMC15 SMC16 SMC17 SMC18 SMC19 SMC20 SMC21 SMC22 SMC23
## Malignant cells  9.65 17.96 92.14 38.07 78.40 24.82 35.71 73.75 81.02 39.41
## Macrophages     30.82  6.84  1.39 25.08  3.47  1.38 20.29  2.36  3.28 12.60
## cDC             2.29  1.08  0.03  1.29  0.47  0.39  1.58  0.13  0.30  1.42
## B cells         2.18 15.51  0.27  0.84  1.86 15.26  2.40  2.19  0.77  1.91
## Plasma cells     1.24  6.04  0.14  2.43  1.29  6.29  3.70  0.69  0.21  5.31
## T cells         36.35 35.12  5.05 16.87 14.14 49.44 10.69 20.72 13.74 37.31
## Fibroblasts      8.06  6.56  0.54  5.78  0.10  1.57 20.90  0.17  0.17  0.86
## Stromal cells    3.76  5.25  0.14  4.48  0.00  0.46  3.15  0.00  0.09  0.37
## Endothelial cells 5.65  5.64  0.31  5.17  0.26  0.39  1.58  0.00  0.43  0.80
## SMC24 SMC25
## Malignant cells  20.52 25.45
## Macrophages     21.89  3.64
## cDC             1.50  0.80
## B cells         8.14 10.13
## Plasma cells     2.35  5.56
## T cells         42.02 52.75
## Fibroblasts      1.82  0.25
## Stromal cells    0.52  0.12
## Endothelial cells 1.24  1.30
```

```
prop_ct_plot <- data.frame(Cell_type = rownames(prop_ct), prop_ct)
prop_ct_plot <- melt(prop_ct_plot)
```

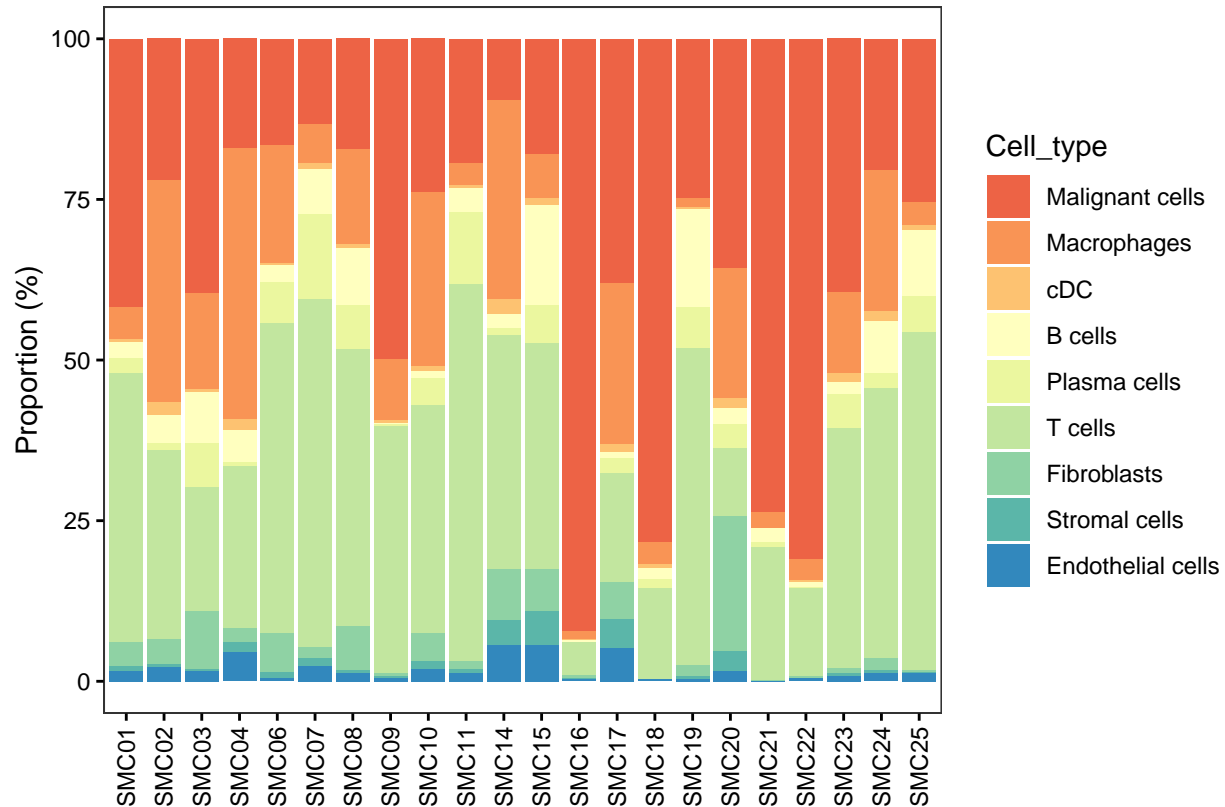
```
## Using Cell_type as id variables
```

```
prop_ct_plot$Cell_type <- factor(prop_ct_plot$Cell_type, levels = c("Malignant cells", "Macrophages", "cDC", "B cells", "Plasma cells", "T cells", "Fibroblasts", "Stromal cells", "Endothelial cells"))
head(prop_ct_plot)
```

```
##      Cell_type variable value
## 1 Malignant cells   SMC01 41.80
## 2   Macrophages     SMC01  5.01
## 3         cDC       SMC01  0.35
## 4        B cells     SMC01  2.49
## 5   Plasma cells     SMC01  2.45
## 6        T cells     SMC01 41.92
```

```
ggplot(prop_ct_plot, aes(variable, value, fill = Cell_type)) +
  geom_bar(stat = 'identity', position = 'stack') +
  scale_fill_manual(values = c("Malignant cells" = "#ED6345", "Macrophages" = "#F99455", "cDC" = "#FDC2F2", "B cells" = "#FFFFBF", "Plasma cells" = "#EBF79F", "T cells" = "#C2E69F", "Fibroblasts" = "#8FD2A4", "Stromal cells" = "#5BB6A9", "Endothelial cells" = "#4DAF4A")) +
  labs(x = '', y = 'Proportion (%)') +
  theme_bw() +
```

```
theme(panel.grid = element_blank(),
      axis.text = element_text(color = 'black'),
      axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5),
      axis.ticks = element_line(color = 'black'))
```



```
#ggsave('../res_various/res_0.3/cluster_proportion/prop.celltype.tumor.pdf', unit = 'cm', width = 14, height = 10)
```

```
### Correlation between malignant clusters and cell types
```

```
library(pheatmap)
```

```
cor(t(prop), t(prop_ct), method = 'pearson')
```

```
##      Malignant cells  Macrophages      cDC      B cells Plasma cells
## C0      0.19329773 -0.241594669 -0.44795627  0.09290279 -0.01522599
## C1      0.09005550  0.326518710  0.25788993 -0.30723813 -0.35284229
## C2      0.48698735 -0.395377753 -0.52120583 -0.35718677 -0.19444591
## C3     -0.51427934  0.243613391  0.36327769  0.18410472  0.44250640
## C4      0.06892809 -0.007503062  0.14361236  0.13181509 -0.07919651
## C5     -0.19045112  0.151000327  0.40979767  0.03138115  0.04988788
## C6     -0.27578108 -0.060224299  0.02991752  0.40561974  0.19204616
##      T cells Fibroblasts Stromal cells Endothelial cells
## C0  0.02149703 -0.09940740  -0.50801418  -0.54047426
## C1 -0.43110337  0.28973000   0.50034378   0.51477981
## C2 -0.09257609 -0.39483860  -0.37849339  -0.31497250
```

```
## C3  0.41427754  0.02043477    0.02548592    0.21043556
## C4 -0.20978166  0.22439177    0.18429637   -0.04826474
## C5  0.11617501 -0.01755703    0.15021603    0.16119317
## C6  0.21807756  0.04785162    0.47527471    0.31219196
```

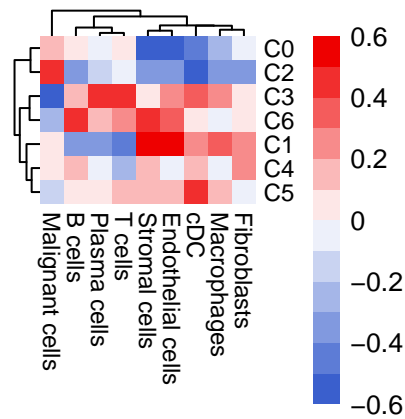
```
min(cor(t(prop), t(prop_ct), method = 'pearson'))
```

```
## [1] -0.5404743
```

```
max(cor(t(prop), t(prop_ct), method = 'pearson'))
```

```
## [1] 0.5147798
```

```
pheatmap(cor(t(prop), t(prop_ct), method = 'pearson'), #filename = '../res_various/res_0.3/cluster_prop
  treeheight_row = 5, treeheight_col = 5, border_color = NA,
  fontsize_col = 9, fontsize_row = 9, cellwidth = 9, cellheight = 9,
  breaks = c(-0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6),
  color = colorRampPalette(c("#3a5fcd", 'white', "#ee0000"))(n = 12))
```



```
#dev.off()
```