First AACES analysis

Ariel Hippen

2023-01-28

Contents

Having run the TCGA RNA-seq data through BayesPrism, this notebook compares the samples' cell type composition with their subtype annotations from the Way pipeline and the patients' survival status/time.

```r
suppressPackageStartupMessages({
    library(data.table)
    library(SingleCellExperiment)
    library(dplyr)
    library(yaml)
    library(stringr)
    library(ggplot2)
    library(survival)
    library(ggfortify)
})

params <- read_yaml("../../config.yml")
data_path <- params$data_path
local_data_path <- params$local_data_path
plot_path <- params$plot_path
```

```r
aaces <- fread(paste(local_data_path, "deconvolution_output",
                     "AACES_default_bayesprism_results.tsv", sep = "/"))
aaces_melt <- melt(aaces)
```

```
## Warning in melt.data.table(aaces): id.vars and measure.vars are internally
## guessed when both are 'NULL'. All non-numeric/integer/logical type columns are
## considered id.vars, which in this case are columns [cell_type, ...]. Consider
## providing at least one of 'id' or 'measure' vars in future.
```

Cell composition

```r
g <- ggplot(aaces_melt, mapping = aes(x=variable, y=value, fill=cell_type, color=cell_type)) +
    geom_bar(stat = "identity") +
    theme(axis.text.x=element_blank(), #remove x axis labels
          axis.ticks.x=element_blank(), #remove x axis ticks
          axis.text.y=element_blank(),  #remove y axis labels
          axis.ticks.y=element_blank()) #remove y axis ticks

g
```
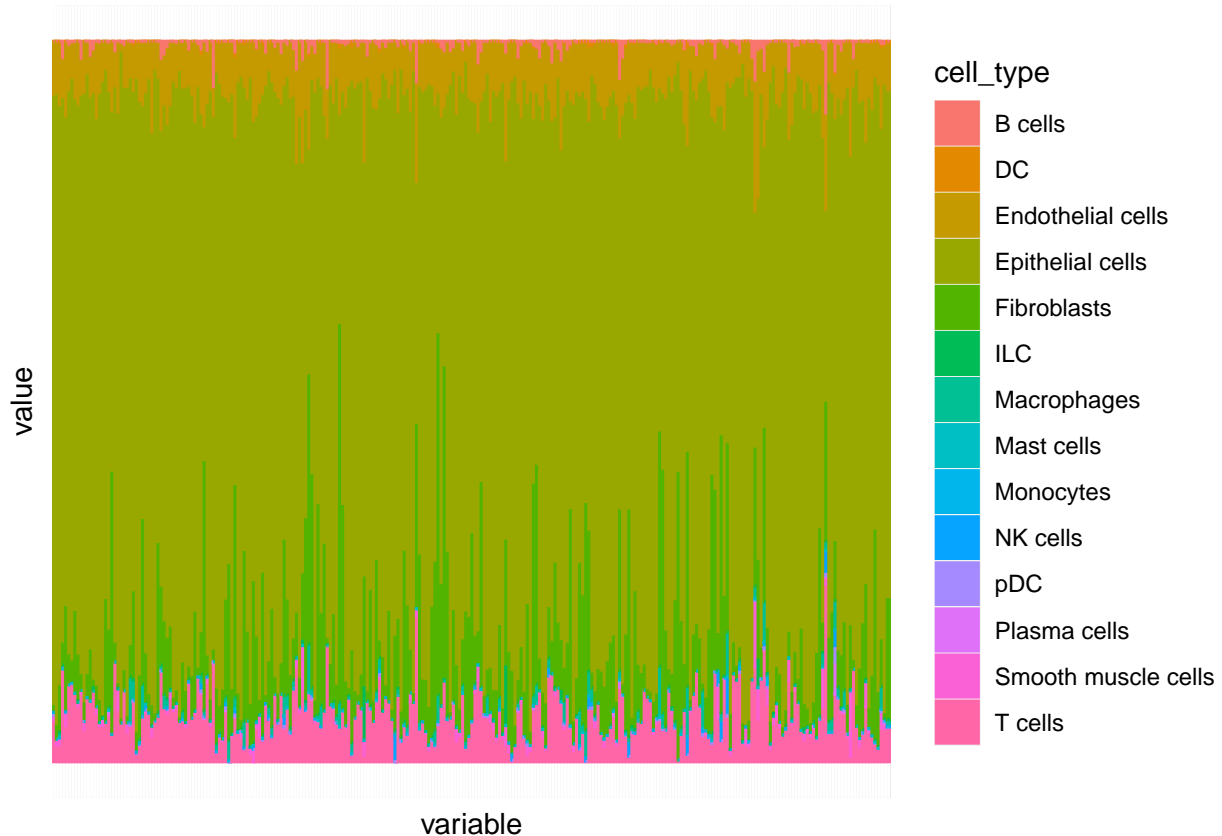


```r
plotfile <- paste(plot_path, "evaluation_plots",
                  "AACES_proportion_barchart.png", sep = "/")
png(filename = plotfile, width = 1200)
g
dev.off()
```

```
## pdf
##   2
```

```r
# Switch so cell types are columns and samples are rows for easier analysis
cell_types <- aaces$cell_type

aaces$cell_type <- NULL
aaces_t <- t(as.matrix(aaces))
colnames(aaces_t) <- cell_types
```

```r
aaces_t <- as.data.frame(aaces_t)
aaces_t <- cbind(rownames(aaces_t), aaces_t)
setnames(aaces_t, "rownames(aaces_t)", "ID")
```

Cell composition by survival

```r
# Load survival data
aaces_survival <- fread(paste(local_data_path, "AACES", "AACES_survival.tsv",
                              sep = "/"))
aaces_t <- left_join(aaces_t, aaces_survival)
```

```
## Joining, by = "ID"
```

```r
aaces_t$Immune <- aaces_t$Macrophages + aaces_t$Monocytes + aaces_t$`Plasma cells` +
  aaces_t$DC + aaces_t$`NK cells` + aaces_t$pDC + aaces_t$`B cells` + aaces_t$ILC +
  aaces_t$`Mast cells`
```

```r
g <- ggplot(aaces_t, mapping = aes(x=survival_days, y=Fibroblasts)) +
  geom_point() + xlab("OS (days)") + ylab("% Fibroblasts")
plotfile <- paste(plot_path, "evaluation_plots", "AACES_survival_by_fibroblasts.png", sep = "/")
png(filename = plotfile); g; dev.off()
```
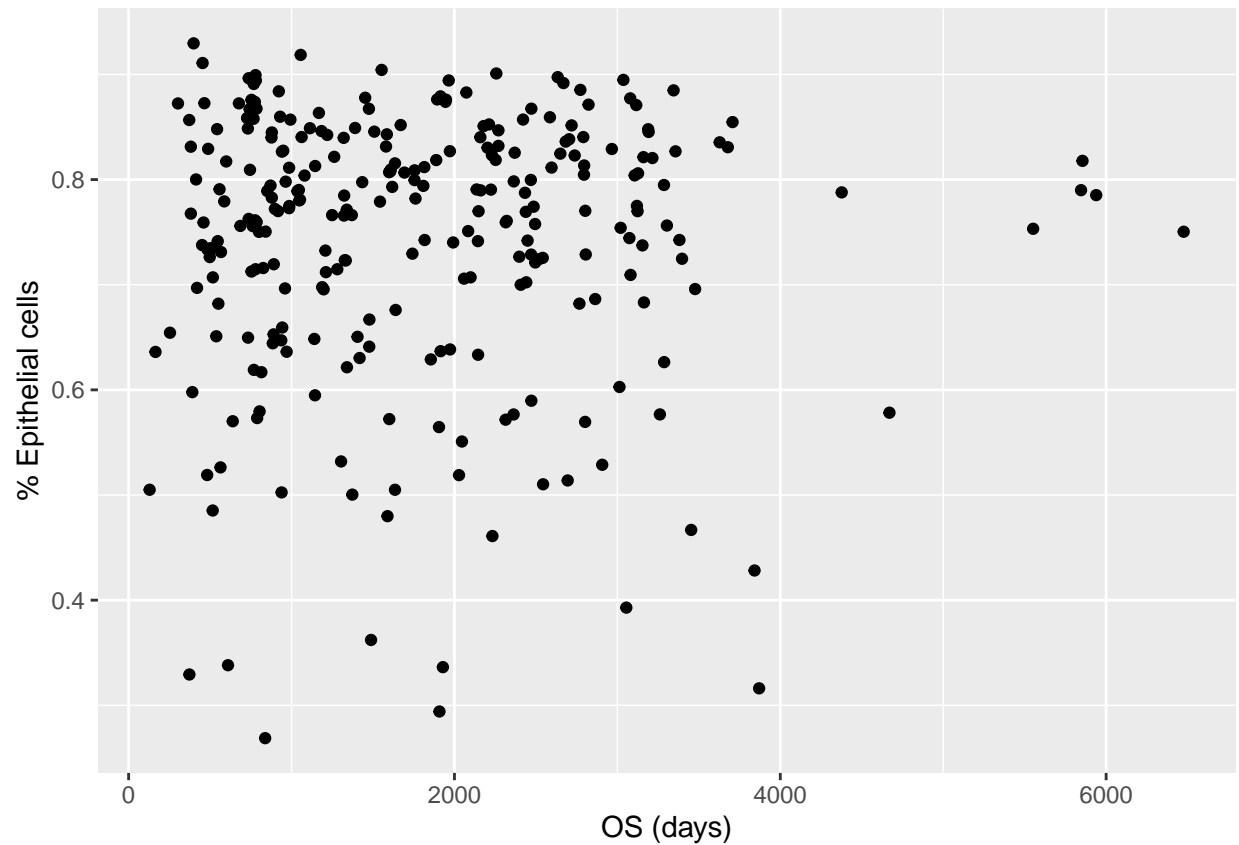
```
## pdf
##   2
```

```r
g
```
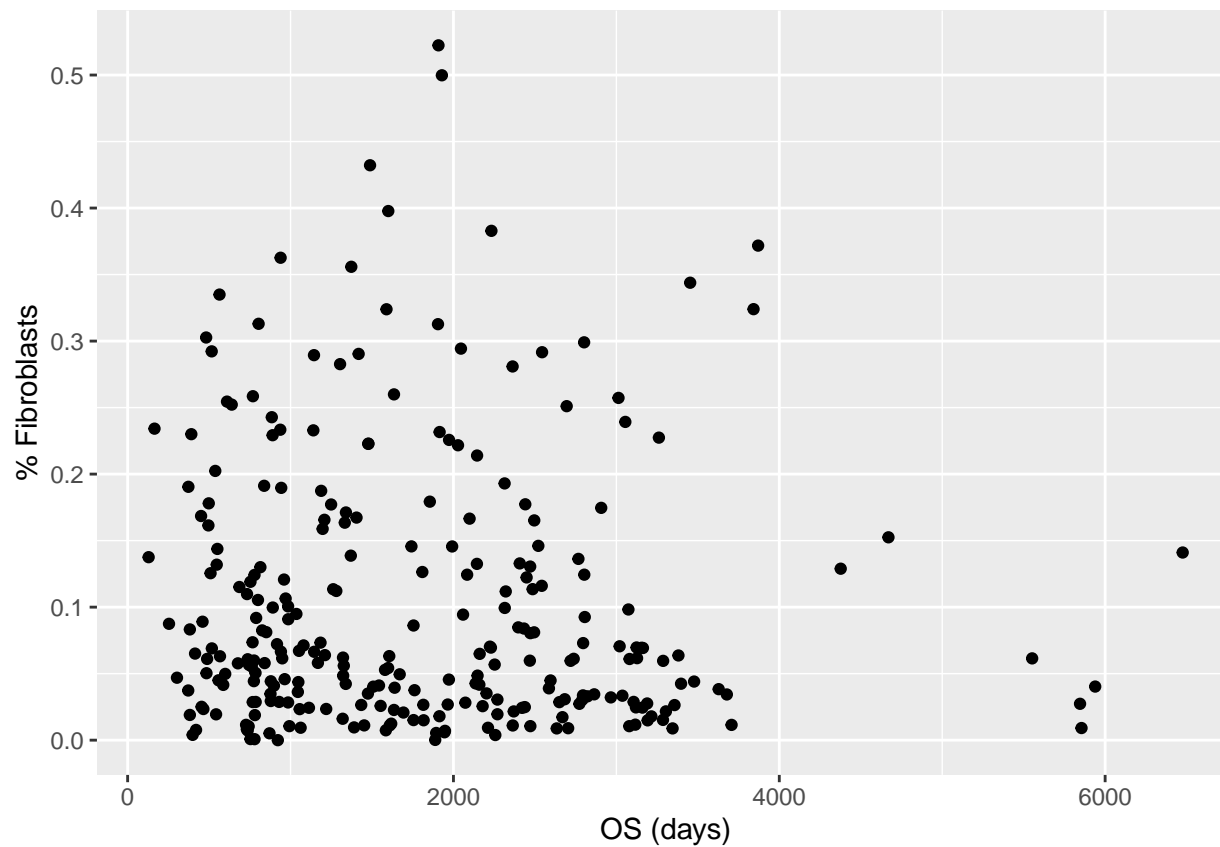
```
g <- ggplot(aaces_t, mapping = aes(x=survival_days, y=`Epithelial cells`)) +
  geom_point() + xlab("OS (days)") + ylab("% Epithelial cells")
plotfile <- paste(plot_path, "evaluation_plots", "AACES_survival_by_epithelial.png", sep = "/")
png(filename = plotfile); g; dev.off()
```
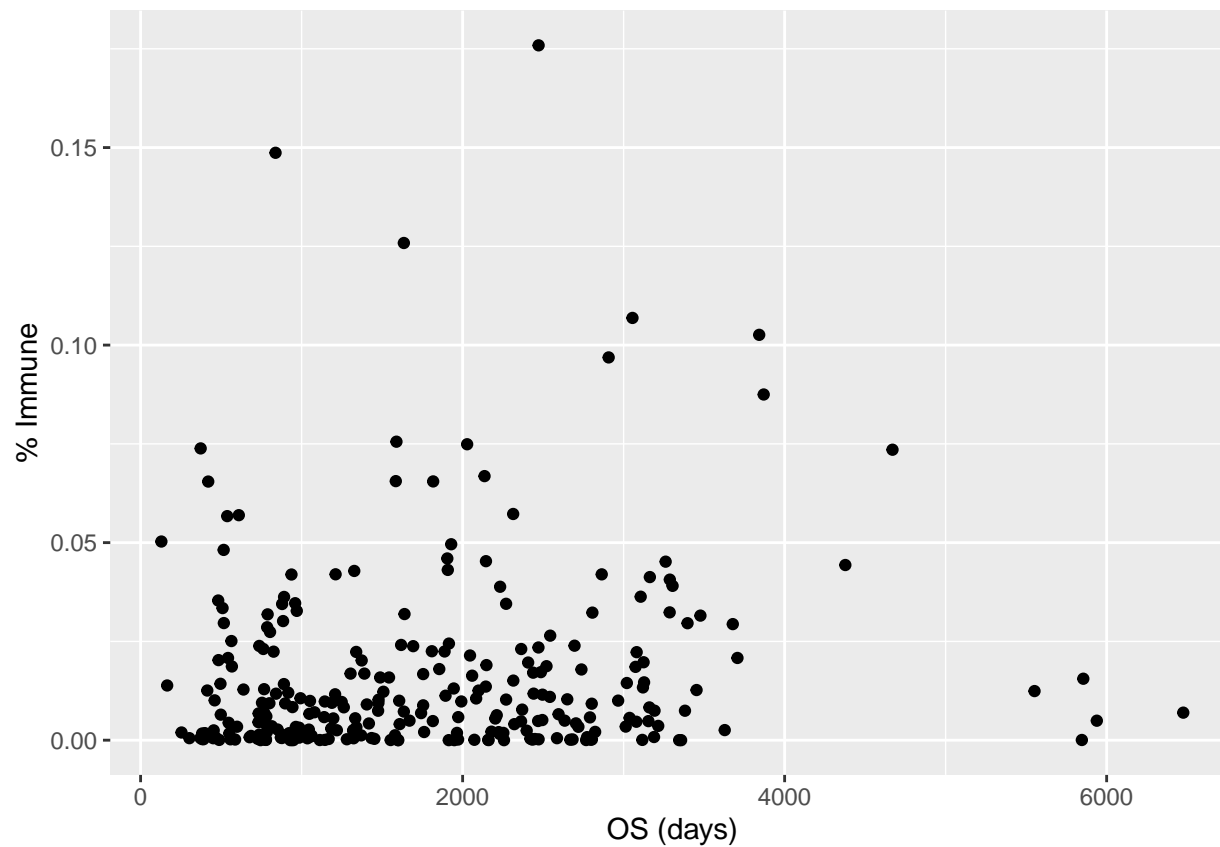
```
## pdf
##   2
```

```
g
```

```
g <- ggplot(aaces_t, mapping = aes(x=survival_days, y=Immune)) +
  geom_point() + xlab("OS (days)") + ylab("% Immune")
plotfile <- paste(plot_path, "evaluation_plots", "AACES_survival_by_immune.png", sep = "/")
png(filename = plotfile); g; dev.off()
```
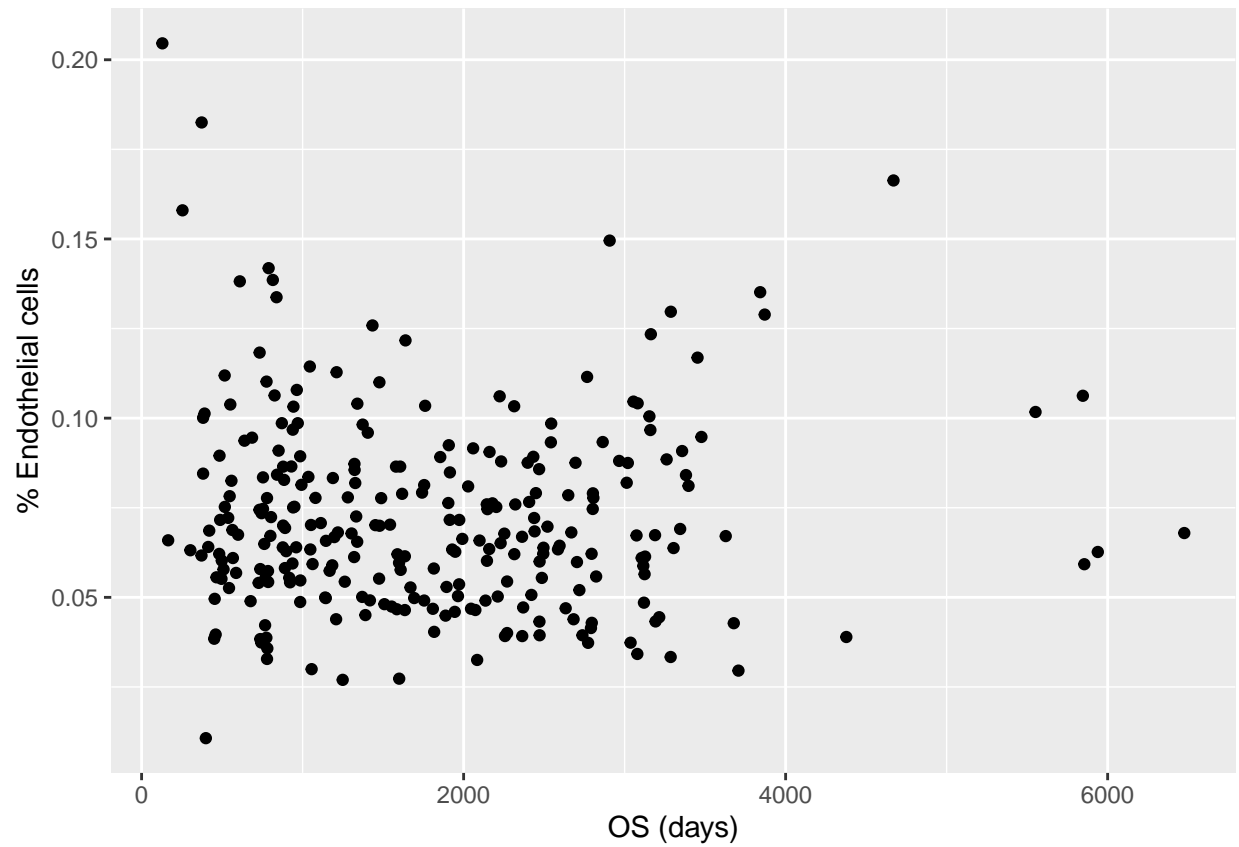
```
## pdf
##   2
```

```
g
```

```
g <- ggplot(aaces_t, mapping = aes(x=survival_days, y=`Endothelial cells`)) +
  geom_point() + xlab("OS (days)") + ylab("% Endothelial cells")
plotfile <- paste(plot_path, "evaluation_plots", "AACES_survival_by_endothelial_cells.png", sep = "/")
png(filename = plotfile); g; dev.off()
```
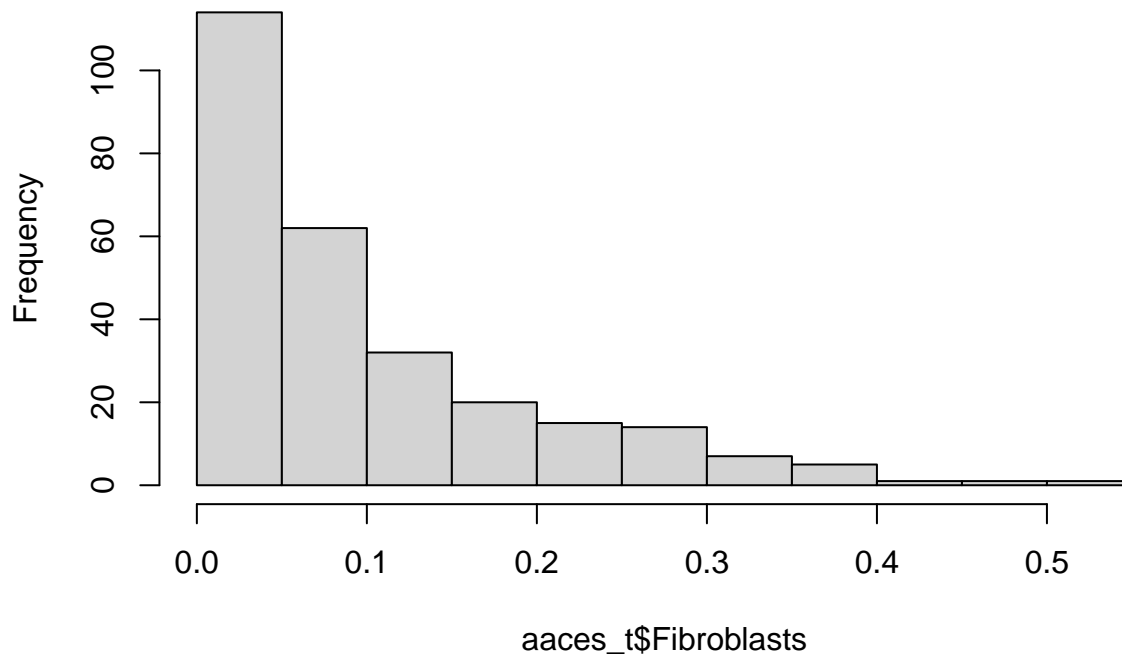
```
## pdf
##   2
```

```
g
```

Kaplan Meier curves

```
# Put the samples into quartiles based on fibroblast content
hist(aaces_t$Fibroblasts)
```

Histogram of aaces_t$Fibroblasts


```r
quantiles <- quantile(aaces_t$Fibroblasts)
q1 <- quantiles[2]
q3 <- quantiles[4]
aaces_t$high_fibro <- ifelse(aaces_t$Fibroblasts > q3, 1, 0)
```
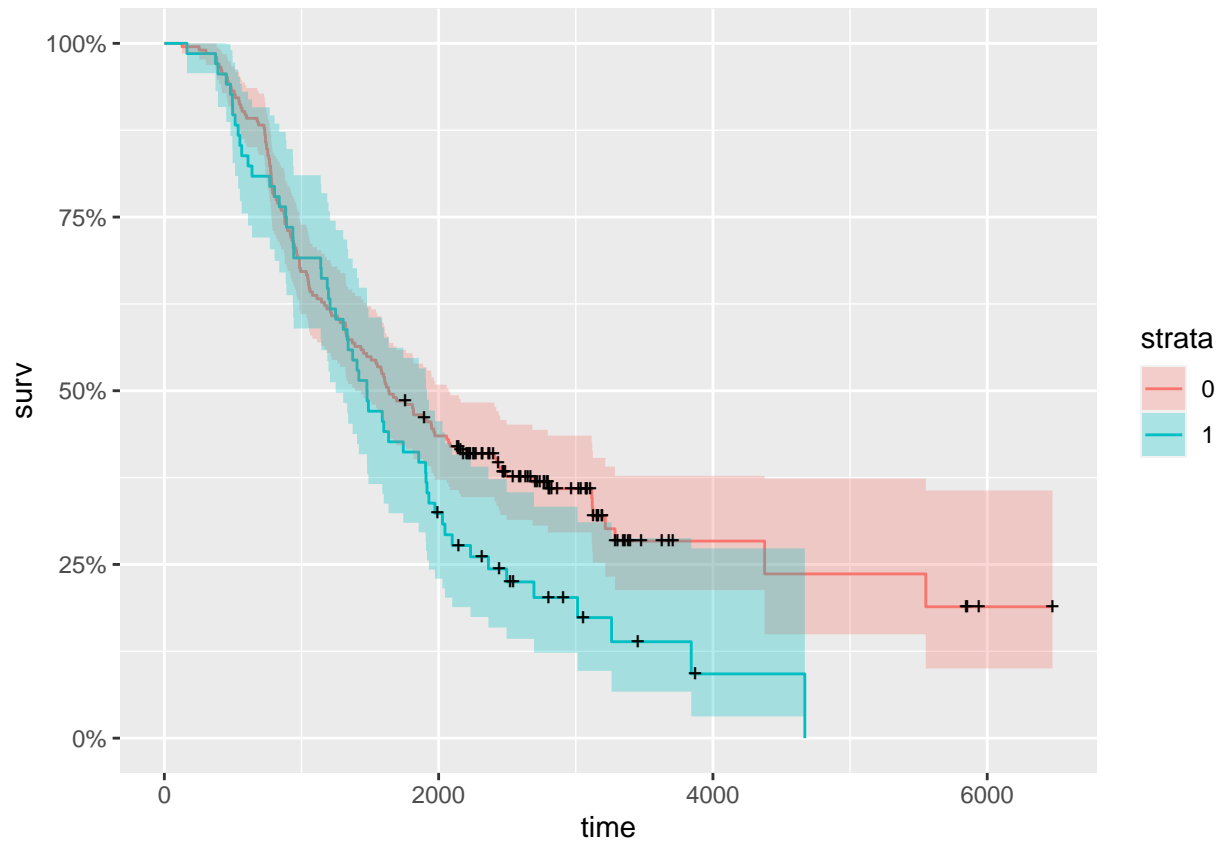
```r
# Get Kaplan-Meier curves
aaces_t$vitalstatus <- ifelse(aaces_t$vitalstatus=="Alive", 0, 1)
km <- Surv(time = aaces_t$survival_days,  event = aaces_t$vitalstatus)
km_treatment<-survfit(km~high_fibro,data=aaces_t,type='kaplan-meier',conf.type='log')

autoplot(km_treatment)
```
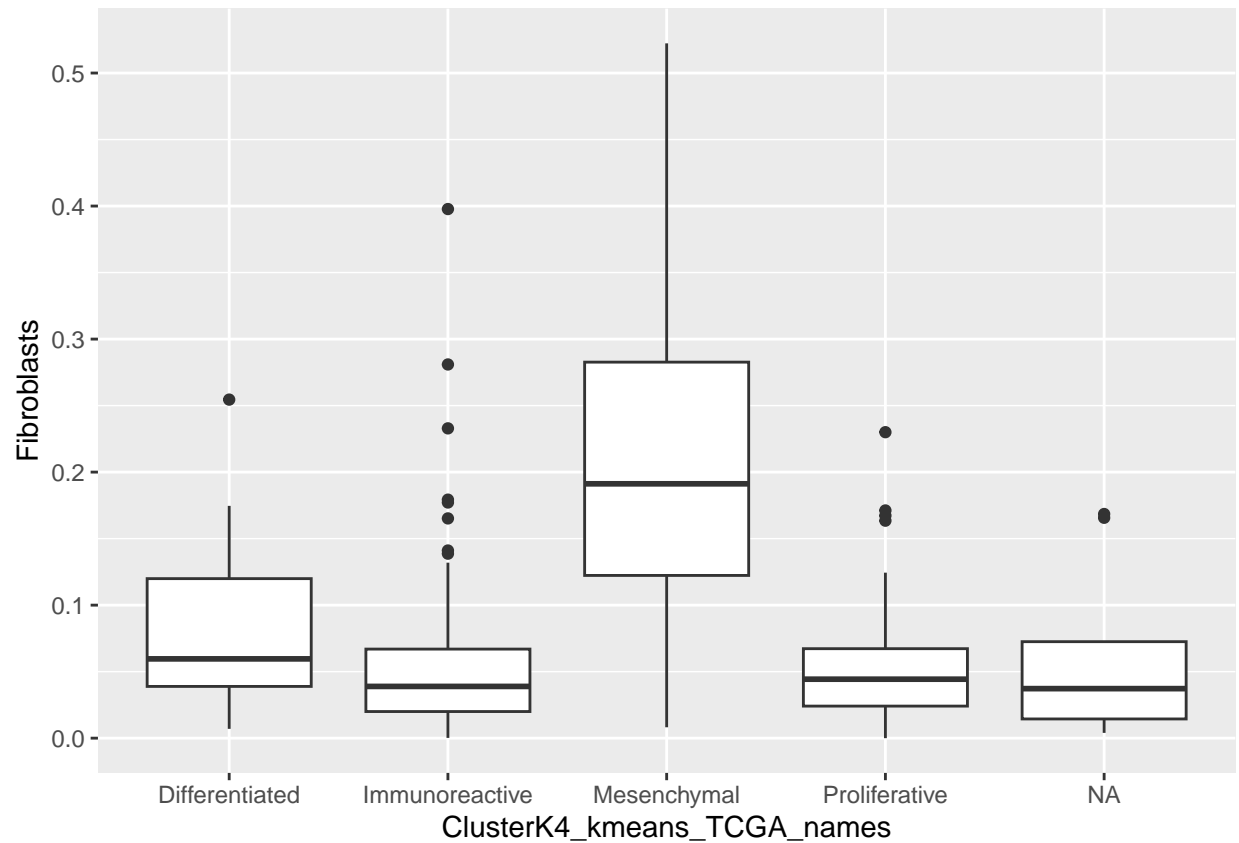
Subtypes

```r
# Compare cell type proportions of subtypes
g <- ggplot(aaces_t, mapping = aes(x=ClusterK4_kmeans_TCGA_names, y=Fibroblasts)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "AACES_fibroblasts_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```
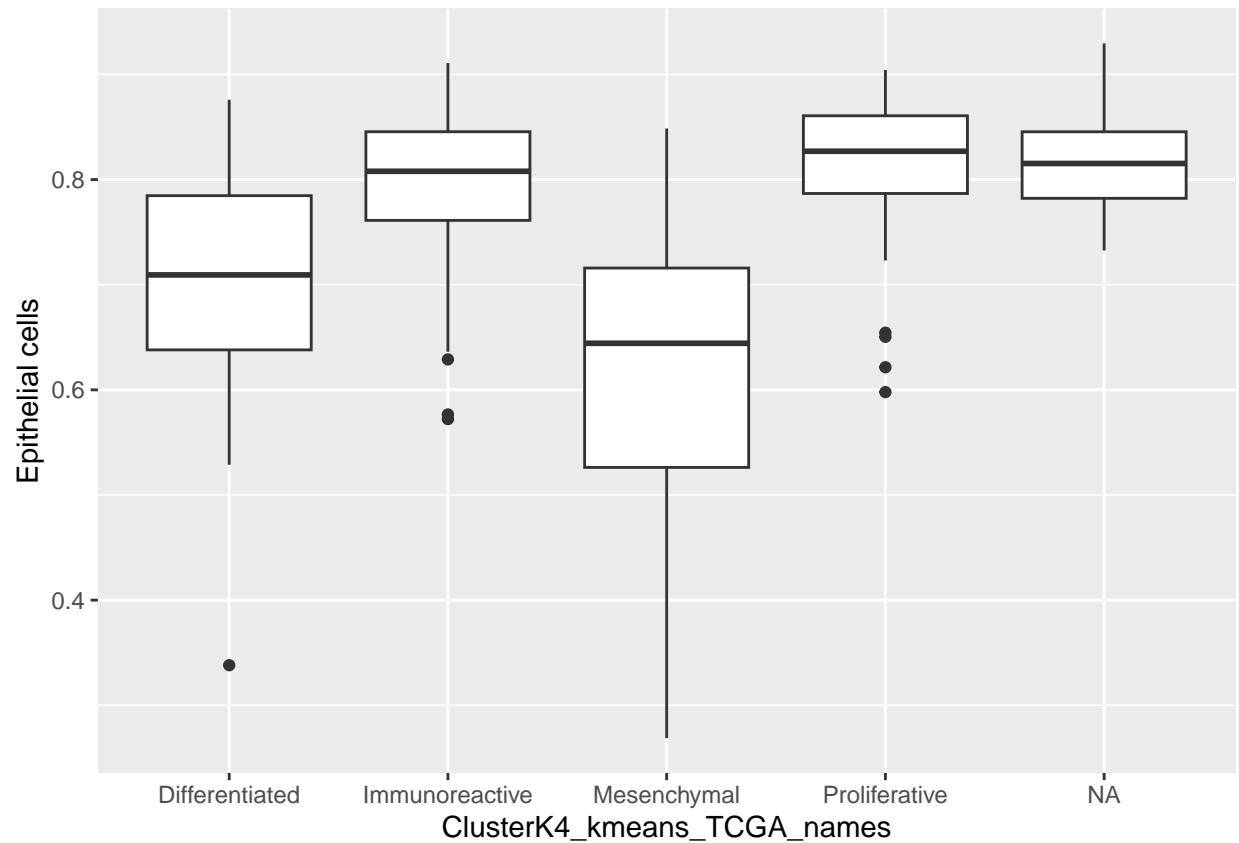
```
## pdf
##   2
```

```
g
```

```
g <- ggplot(aaces_t, mapping = aes(x=ClusterK4_kmeans_TCGA_names, y=`Epithelial cells`)) + geom_boxplot
plotfile <- paste(plot_path, "evaluation_plots", "AACES_epithelial_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```
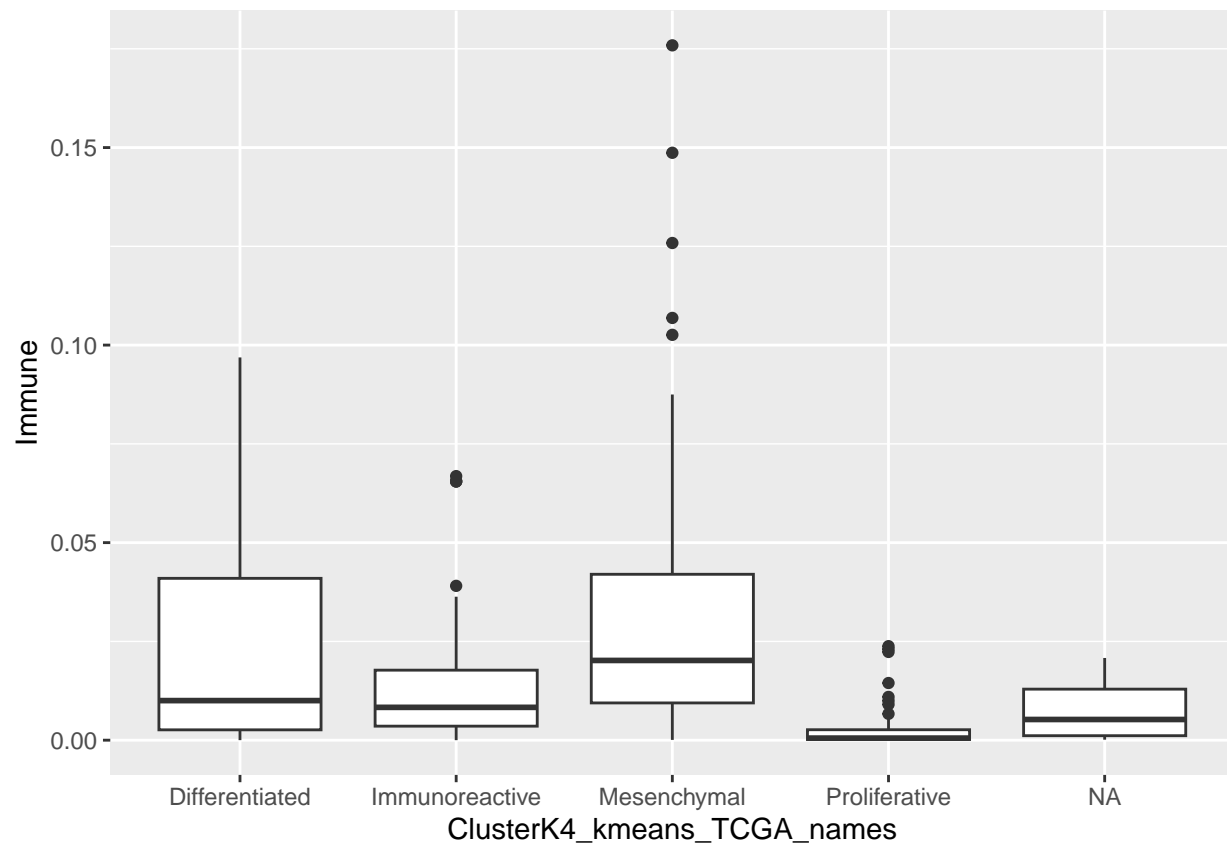
```
## pdf
##   2
```

```
g
```

```
g <- ggplot(aaces_t, mapping = aes(x=ClusterK4_kmeans_TCGA_names, y=Immune)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "AACES_immune_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

```
g
```

11

```
g <- ggplot(aaces_t, mapping = aes(x=ClusterK4_kmeans_TCGA_names, y=`Endothelial cells`)) + geom_boxplo
plotfile <- paste(plot_path, "evaluation_plots", "AACES_endothelial_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

```
g
```