

# HGSOC Survival

Ariel Hippen

2023-02-18

## Contents

Load data . . . . .	1
AACES . . . . .	3
TCGA . . . . .	4
Tothill . . . . .	5
All . . . . .	6

In `first_*_analysis.Rmd`, we did a preliminary survival analysis of how fibroblast content affects survival, but with the simplest model possible. Here, we will redo that analysis but considering important covariates like age and tumor stage.

```
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
```

## Load data

```
aaces_t <- fread(paste(local_data_path, "deconvolution_output",
                        "AACES_default_bayesprism_results.tsv", sep = "/"))
tcga_t <- fread(paste(local_data_path, "deconvolution_output",
                        "TCGA_default_bayesprism_results.tsv", sep = "/"))
tothill_t <- fread(paste(local_data_path, "deconvolution_output",
                           "tothill_default_bayesprism_results.tsv", sep = "/"))
```

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

```
aaces_t$cell_type <- NULL
aaces <- t(as.matrix(aaces_t))
colnames(aaces) <- cell_types
aaces <- as.data.frame(aaces)
aaces <- cbind(rownames(aaces), aaces)
setnames(aaces, "rownames(aaces)", "ID")
aaces$dataset <- "AA"
```

```
tcga_t$cell_type <- NULL
tcga <- t(as.matrix(tcga_t))
colnames(tcga) <- cell_types
tcga <- as.data.frame(tcga)
tcga <- cbind(rownames(tcga), tcga)
setnames(tcga, "rownames(tcga)", "ID")
tcga$dataset <- "TCGA"
```

```
tothill_t$cell_type <- NULL
tothill <- t(as.matrix(tothill_t))
colnames(tothill) <- cell_types
tothill <- as.data.frame(tothill)
tothill <- cbind(rownames(tothill), tothill)
setnames(tothill, "rownames(tothill)", "ID")
tothill$dataset <- "tothill"
```

```
rm(aaces_t, tcga_t, tothill_t); gc()
```

```
##          used (Mb) gc trigger (Mb) max used (Mb)
## Ncells  6044576 322.9   11849142 632.9   7655865 408.9
## Vcells 10595812  80.9    18479016 141.0  12706855  97.0
```

```
tcga$ID <- str_extract(tcga$ID, "TCGA-\\w\\w-\\w\\w\\w\\w\\w")
tcga$ID <- gsub("-", "\\.", tcga$ID)
```

```
# Combine cell type info into one object to run combined survival analysis
composition <- rbind(aaces, tcga, tothill)
```

```
covariates <- fread(paste(local_data_path, "cluster_assignments",
                          "AnalSet.csv", sep = "/"))
covariates$V1 <- NULL
setnames(covariates, "sampleid", "ID")

covariates$debulking <- recode(covariates$debulking,
                              "Optimal" = "optimal",
                              "Suboptimal" = "suboptimal")

composition <- inner_join(composition, covariates)
```

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

## AACES

```
aaces <- subset(composition, composition$Dataset=="AA")
```

```
# Discretize fibroblast estimates
quantiles <- quantile(aaces$Fibroblasts)
q1 <- quantiles[2]
median <- quantiles[3]
q3 <- quantiles[4]
aaces$high_fibro <- ifelse(aaces$Fibroblasts > q3, 1, 0)
```

```
# Basic analysis
model <- coxph(Surv(months, vital) ~ (factor(high_fibro)), aaces)
model
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro)), data = aaces)
##
##               coef exp(coef) se(coef)      z      p
## factor(high_fibro)1 0.3164    1.3721   0.1617 1.956 0.0504
##
## Likelihood ratio test=3.66 on 1 df, p=0.05569
## n= 262, number of events= 183
```

```
# Analysis with covariates
model <- coxph(Surv(months, vital) ~ (factor(high_fibro)+ factor(age)+
                                     factor(debulking)+ factor(FewerStage)), aaces)
model
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro) + factor(age) +
##   factor(debulking) + factor(FewerStage)), data = aaces)
##
##               coef exp(coef) se(coef)      z      p
## factor(high_fibro)1    0.2181    1.2437   0.1737  1.256 0.209105
## factor(age)2          0.2690    1.3087   0.6630  0.406 0.684939
## factor(age)3          0.7638    2.1464   0.6326  1.207 0.227334
## factor(age)4          0.7546    2.1268   0.6217  1.214 0.224797
## factor(age)5          0.7819    2.1857   0.6124  1.277 0.201633
## factor(age)6          0.8731    2.3943   0.6147  1.420 0.155471
## factor(age)7          0.8792    2.4090   0.6278  1.401 0.161351
## factor(age)8          1.1331    3.1052   0.6330  1.790 0.073433
## factor(age)9          0.4655    1.5928   0.7207  0.646 0.518360
## factor(debulking)optimal -0.6549    0.5195   0.1829 -3.581 0.000342
## factor(debulking)suboptimal 0.1114    1.1178   0.1938  0.575 0.565557
## factor(FewerStage)2    1.2587    3.5208   0.3362  3.744 0.000181
## factor(FewerStage)4    1.3562    3.8814   0.3922  3.458 0.000544
## factor(FewerStage)9    0.9030    2.4671   0.4272  2.114 0.034539
##
## Likelihood ratio test=55.87 on 14 df, p=6.134e-07
## n= 262, number of events= 183
```

## TCGA

```
tcga <- subset(composition, composition$Dataset=="TCGA")
```

```
# Discretize fibroblast estimates
quantiles <- quantile(tcga$Fibroblasts)
q1 <- quantiles[2]
median <- quantiles[3]
q3 <- quantiles[4]
tcga$high_fibro <- ifelse(tcga$Fibroblasts > q3, 1, 0)
```

```
# Basic analysis
model <- coxph(Surv(months, vital) ~ (factor(high_fibro)), tcga)
model
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro)), data = tcga)
##
##               coef exp(coef) se(coef)      z    p
## factor(high_fibro)1 0.3052    1.3570   0.1857 1.644 0.1
##
## Likelihood ratio test=2.58 on 1 df, p=0.108
## n= 268, number of events= 149
## (5 observations deleted due to missingness)
```

```
# Analysis with covariates
model <- coxph(Surv(months, vital) ~ (factor(high_fibro)+ factor(age)+
                                     factor(debulking)+ factor(FewerStage)), tcga)
model
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro) + factor(age) +
##   factor(debulking) + factor(FewerStage)), data = tcga)
##
##               coef exp(coef) se(coef)      z    p
## factor(high_fibro)1    0.03966    1.04046   0.20480 0.194 0.846
## factor(age)2           0.39555    1.48520   1.10604 0.358 0.721
## factor(age)3           0.39980    1.49153   1.05799 0.378 0.706
## factor(age)4           0.43512    1.54515   1.03708 0.420 0.675
## factor(age)5           0.84831    2.33570   1.03237 0.822 0.411
## factor(age)6           0.58600    1.79679   1.05275 0.557 0.578
## factor(age)7           0.70352    2.02085   1.03529 0.680 0.497
## factor(age)8           0.98423    2.67576   1.03765 0.949 0.343
## factor(age)9           1.40056    4.05746   1.04168 1.345 0.179
## factor(debulking)suboptimal 0.10506    1.11078   0.19701 0.533 0.594
## factor(FewerStage)4     0.13680    1.14660   0.28576 0.479 0.632
##
## Likelihood ratio test=13.39 on 11 df, p=0.2685
## n= 241, number of events= 136
## (32 observations deleted due to missingness)
```

## Tothill

```
tothill <- subset(composition, composition$Dataset=="Tothill")
```

```
# Discretize fibroblast estimates
```

```
quantiles <- quantile(tothill$Fibroblasts)
q1 <- quantiles[2]
median <- quantiles[3]
q3 <- quantiles[4]
tothill$high_fibro <- ifelse(tothill$Fibroblasts > q3, 1, 0)
```

```
# Basic analysis
```

```
model <- coxph(Surv(months, vital) ~ (factor(high_fibro)), tothill)
model
```

```
## Call:
```

```
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro)), data = tothill)
```

```
##
```

```
##               coef exp(coef) se(coef)      z      p
```

```
## factor(high_fibro)1 0.4540    1.5746    0.2544 1.784 0.0743
```

```
##
```

```
## Likelihood ratio test=3 on 1 df, p=0.08313
```

```
## n= 145, number of events= 73
```

```
## (2 observations deleted due to missingness)
```

```
# Analysis with covariates
```

```
model <- coxph(Surv(months, vital) ~ (factor(high_fibro)+ factor(age)+
                                         factor(debulking)+factor(FewerStage)), tothill)
model
```

```
## Call:
```

```
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro) + factor(age) +
```

```
##      factor(debulking) + factor(FewerStage)), data = tothill)
```

```
##
```

```
##               coef exp(coef) se(coef)      z      p
```

```
## factor(high_fibro)1      0.5407    1.7172    0.2859  1.891 0.0586
```

```
## factor(age)2      -0.9588    0.3833    1.2025 -0.797 0.4252
```

```
## factor(age)3      0.1711    1.1866    0.7698  0.222 0.8241
```

```
## factor(age)4     -1.1170    0.3273    0.7490 -1.491 0.1359
```

```
## factor(age)5     -0.6915    0.5008    0.6663 -1.038 0.2994
```

```
## factor(age)6     -0.1862    0.8301    0.7159 -0.260 0.7948
```

```
## factor(age)7     -1.4104    0.2440    0.8101 -1.741 0.0817
```

```
## factor(age)8      0.7986    2.2225    0.6966  1.147 0.2516
```

```
## factor(age)9     -0.2931    0.7459    0.7816 -0.375 0.7076
```

```
## factor(debulking)suboptimal -0.4039    0.6677    0.3196 -1.264 0.2064
```

```
## factor(FewerStage)2    0.3426    1.4086    1.0343  0.331 0.7405
```

```
## factor(FewerStage)4    2.2664    9.6446    1.1361  1.995 0.0461
```

```
##
```

```
## Likelihood ratio test=33.54 on 12 df, p=0.0007965
```

```
## n= 128, number of events= 65
```

```
## (19 observations deleted due to missingness)
```

All

```
# Discretize fibroblast estimates
quantiles <- quantile(composition$Fibroblasts)
q1 <- quantiles[2]
median <- quantiles[3]
q3 <- quantiles[4]
composition$high_fibro <- ifelse(composition$Fibroblasts > q3, 1, 0)

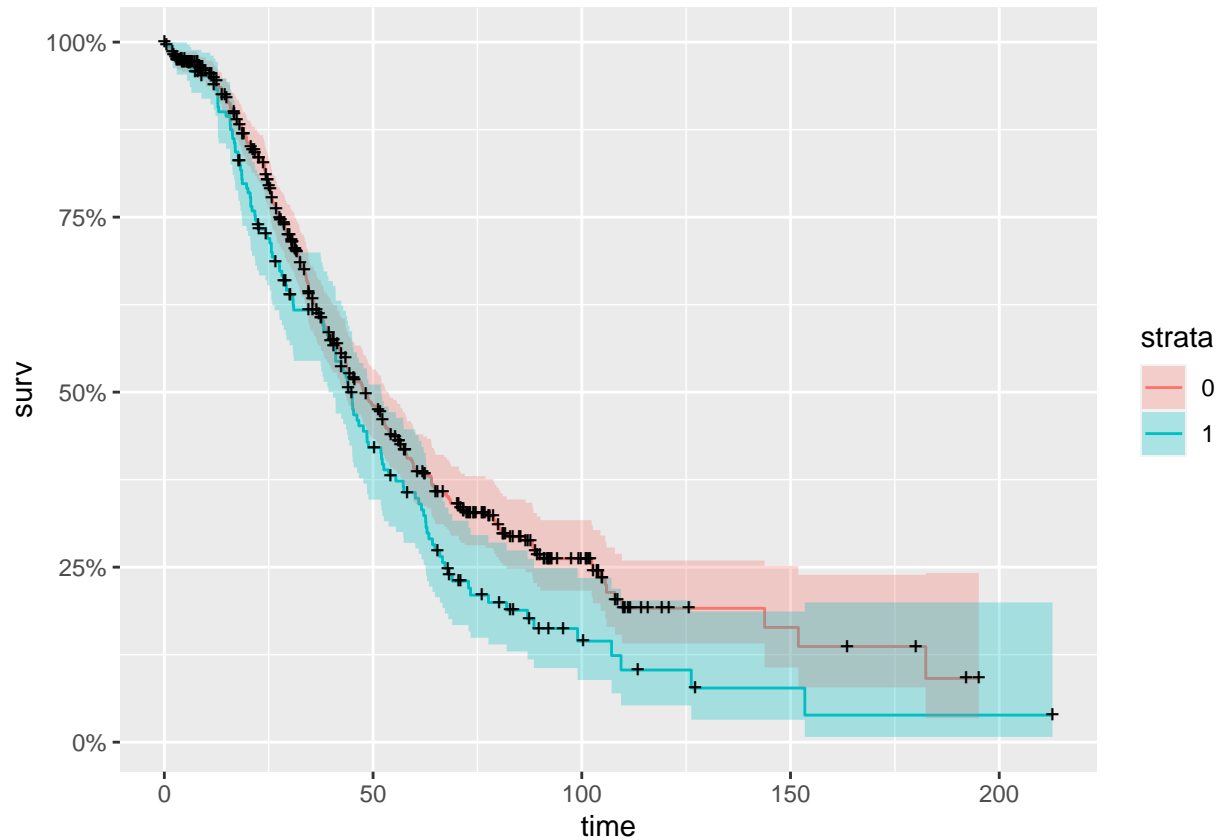
# Basic analysis
model <- coxph(Surv(months, vital) ~ (factor(high_fibro)), composition)
model

## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro)), data = composition)
##
##               coef exp(coef) se(coef)      z      p
## factor(high_fibro)1 0.2486      1.2823   0.1092 2.277 0.0228
##
## Likelihood ratio test=5   on 1 df, p=0.0253
## n= 675, number of events= 405
##      (7 observations deleted due to missingness)

# Analysis with covariates
model <- coxph(Surv(months, vital) ~ (factor(high_fibro)+ factor(Dataset)+ factor(age)+
                                     factor(debulking) + factor(FewerStage)), composition)
model

## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro) + factor(Dataset) +
##      factor(age) + factor(debulking) + factor(FewerStage)), data = composition)
##
##               coef exp(coef) se(coef)      z      p
## factor(high_fibro)1      0.16272   1.17671  0.11623  1.400 0.161517
## factor(Dataset)TCGA      0.47780   1.61253  0.13169  3.628 0.000285
## factor(Dataset)Tothill    0.47312   1.60499  0.16311  2.901 0.003724
## factor(age)2              0.04361   1.04457  0.45390  0.096 0.923459
## factor(age)3              0.43127   1.53921  0.41864  1.030 0.302927
## factor(age)4              0.32894   1.38950  0.40551  0.811 0.417263
## factor(age)5              0.45431   1.57508  0.39788  1.142 0.253525
## factor(age)6              0.63222   1.88179  0.40429  1.564 0.117870
## factor(age)7              0.46011   1.58425  0.41012  1.122 0.261905
## factor(age)8              0.95802   2.60654  0.40819  2.347 0.018924
## factor(age)9              0.84581   2.32986  0.42809  1.976 0.048180
## factor(debulking)optimal  -0.43082   0.64998  0.16352 -2.635 0.008424
## factor(debulking)suboptimal -0.12018   0.88676  0.17527 -0.686 0.492899
## factor(FewerStage)2       1.26428   3.54055  0.31716  3.986 6.71e-05
## factor(FewerStage)4       1.57694   4.84014  0.34409  4.583 4.58e-06
## factor(FewerStage)9       0.88595   2.42528  0.40883  2.167 0.030233
##
## Likelihood ratio test=91.58   on 16 df, p=1.276e-12
## n= 631, number of events= 384
##      (51 observations deleted due to missingness)
```

```
# Kaplan Meier plot for all datasets
km <- Surv(composition$months, composition$vital)
km_treatment<-survfit(km~high_fibro,data=composition,type='kaplan-meier',conf.type='log')
autoplot(km_treatment)
```



Looks like high fibroblast content is associated with worse survival. It doesn't hit statistical significance most of the time, but the results do seem on the same scale as what Mollie saw of the effect of subtypes on survival ([https://github.com/greenelab/hgsc\\_characterization/blob/master/figure\\_notebooks/survival\\_figs/survival\\_analyses.R](https://github.com/greenelab/hgsc_characterization/blob/master/figure_notebooks/survival_figs/survival_analyses.R)). It's obviously not going to matter as much as stage or age, but it's still important to consider.