

HGSOC Survival

Ariel Hippen

2023-02-18

Contents

Load data	1
TCGA microarray	3
TCGA RNA-seq	6
Tothill	8
All	11

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

```
microarray_t <- fread(paste(local_data_path, "deconvolution_output",
                             "microarray_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 <- microarray_t$cell_type
```

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

```
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(microarray_t, tcga_t, tothill_t); gc()
```

```
##          used (Mb) gc trigger (Mb) max used (Mb)
## Ncells  6119068 326.8   9894345 528.5  8392225 448.2
## Vcells 10843886  82.8   18005175 137.4 12255519  93.6
```

```
tcga$ID <- str_extract(tcga$ID, "TCGA-\\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(microarray, tcga, tothill)
```

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

```
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 with 'by = join_by(ID)'
```

TCGA microarray

```
microarray <- subset(composition, composition$dataset=="microarray")
```

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

quantiles <- quantile(microarray$Immune)
q1 <- quantiles[2]
median <- quantiles[3]
q3 <- quantiles[4]
microarray$high_immune <- ifelse(microarray$Immune > q3, 1, 0)
```

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ factor(high_fibro), data = microarray)
##
##      n= 426, number of events= 222
##      (9 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_fibro)1 0.04519   1.04622  0.15496 0.292   0.771
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(high_fibro)1    1.046    0.9558   0.7722   1.417
##
## Concordance= 0.503 (se = 0.017 )
## Likelihood ratio test= 0.08 on 1 df,  p=0.8
## Wald test               = 0.09 on 1 df,  p=0.8
## Score (logrank) test = 0.09 on 1 df,  p=0.8
```

```
model <- coxph(Surv(months, vital) ~ factor(high_immune), microarray)
summary(model)
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ factor(high_immune), data = microarray)
##
##      n= 426, number of events= 222
##      (9 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
```

```
## factor(high_immune)1 0.00193 1.00193 0.15895 0.012 0.99
##
## exp(coef) exp(-coef) lower .95 upper .95
## factor(high_immune)1 1.002 0.9981 0.7337 1.368
##
## Concordance= 0.504 (se = 0.017 )
## Likelihood ratio test= 0 on 1 df, p=1
## Wald test = 0 on 1 df, p=1
## Score (logrank) test = 0 on 1 df, p=1
```

Analysis with covariates

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro) + factor(age) +
## factor(debulking) + factor(FewerStage)), data = microarray)
##
## n= 387, number of events= 207
## (48 observations deleted due to missingness)
##
## coef exp(coef) se(coef) z Pr(>|z|)
## factor(high_fibro)1 -0.03021 0.97024 0.17047 -0.177 0.8593
## factor(age)2 0.28096 1.32440 0.67180 0.418 0.6758
## factor(age)3 0.05322 1.05467 0.64046 0.083 0.9338
## factor(age)4 0.45903 1.58254 0.61052 0.752 0.4521
## factor(age)5 0.64677 1.90937 0.60559 1.068 0.2855
## factor(age)6 0.50431 1.65584 0.61772 0.816 0.4143
## factor(age)7 0.51137 1.66758 0.61513 0.831 0.4058
## factor(age)8 0.82812 2.28901 0.60708 1.364 0.1725
## factor(age)9 1.16007 3.19014 0.61355 1.891 0.0587
## factor(debulking)suboptimal 0.23435 1.26408 0.15584 1.504 0.1326
## factor(FewerStage)2 0.53463 1.70682 1.01291 0.528 0.5976
## factor(FewerStage)4 0.70212 2.01802 1.02879 0.682 0.4949
## factor(FewerStage)9 0.60470 1.83069 1.43138 0.422 0.6727
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coef) exp(-coef) lower .95 upper .95
## factor(high_fibro)1 0.9702 1.0307 0.6947 1.355
## factor(age)2 1.3244 0.7551 0.3550 4.941
## factor(age)3 1.0547 0.9482 0.3006 3.701
## factor(age)4 1.5825 0.6319 0.4783 5.236
## factor(age)5 1.9094 0.5237 0.5827 6.257
## factor(age)6 1.6558 0.6039 0.4934 5.557
## factor(age)7 1.6676 0.5997 0.4994 5.568
## factor(age)8 2.2890 0.4369 0.6965 7.523
## factor(age)9 3.1901 0.3135 0.9584 10.618
## factor(debulking)suboptimal 1.2641 0.7911 0.9314 1.716
## factor(FewerStage)2 1.7068 0.5859 0.2344 12.427
## factor(FewerStage)4 2.0180 0.4955 0.2687 15.158
## factor(FewerStage)9 1.8307 0.5462 0.1107 30.270
##
```

```
## Concordance= 0.636 (se = 0.021 )
## Likelihood ratio test= 23.04 on 13 df, p=0.04
## Wald test = 23.41 on 13 df, p=0.04
## Score (logrank) test = 24.41 on 13 df, p=0.03
```

```
model <- coxph(Surv(months, vital) ~ (factor(high_immune) + factor(age)+
                                     factor(debulking)+ factor(FewerStage)), microarray)
summary(model)
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_immune) +
##     factor(age) + factor(debulking) + factor(FewerStage)), data = microarray)
##
## n= 387, number of events= 207
## (48 observations deleted due to missingness)
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_immune)1      0.18445    1.20256  0.16742  1.102   0.2706
## factor(age)2              0.26896    1.30860  0.67130  0.401   0.6887
## factor(age)3              0.02733    1.02771  0.64026  0.043   0.9660
## factor(age)4              0.44059    1.55363  0.61038  0.722   0.4704
## factor(age)5              0.61842    1.85598  0.60492  1.022   0.3066
## factor(age)6              0.46601    1.59362  0.61732  0.755   0.4503
## factor(age)7              0.45249    1.57221  0.61688  0.734   0.4632
## factor(age)8              0.80602    2.23899  0.60745  1.327   0.1845
## factor(age)9              1.15947    3.18823  0.61215  1.894   0.0582
## factor(debulking)suboptimal 0.22665    1.25440  0.15608  1.452   0.1464
## factor(FewerStage)2       0.51482    1.67334  1.01192  0.509   0.6109
## factor(FewerStage)4       0.70199    2.01776  1.02799  0.683   0.4947
## factor(FewerStage)9       0.63602    1.88895  1.43146  0.444   0.6568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(high_immune)1      1.203    0.8316    0.8662    1.670
## factor(age)2              1.309    0.7642    0.3511    4.878
## factor(age)3              1.028    0.9730    0.2930    3.605
## factor(age)4              1.554    0.6437    0.4697    5.139
## factor(age)5              1.856    0.5388    0.5671    6.074
## factor(age)6              1.594    0.6275    0.4752    5.344
## factor(age)7              1.572    0.6360    0.4693    5.267
## factor(age)8              2.239    0.4466    0.6807    7.364
## factor(age)9              3.188    0.3137    0.9605   10.583
## factor(debulking)suboptimal 1.254    0.7972    0.9238    1.703
## factor(FewerStage)2       1.673    0.5976    0.2303   12.160
## factor(FewerStage)4       2.018    0.4956    0.2691   15.132
## factor(FewerStage)9       1.889    0.5294    0.1142   31.238
##
## Concordance= 0.635 (se = 0.021 )
## Likelihood ratio test= 24.19 on 13 df, p=0.03
## Wald test = 24.63 on 13 df, p=0.03
## Score (logrank) test = 25.67 on 13 df, p=0.02
```

TCGA RNA-seq

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

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

```
quantiles <- quantile(tcga$Immune)
q1 <- quantiles[2]
median <- quantiles[3]
q3 <- quantiles[4]
tcga$high_immune <- ifelse(tcga$Immune > q3, 1, 0)
```

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ factor(high_fibro), data = tcga)
##
##      n= 268, number of events= 149
##      (5 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_fibro)1 0.3052      1.3570   0.1857 1.644      0.1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(high_fibro)1      1.357      0.7369      0.943      1.953
##
## Concordance= 0.525  (se = 0.021 )
## Likelihood ratio test= 2.58  on 1 df,   p=0.1
## Wald test               = 2.7  on 1 df,   p=0.1
## Score (logrank) test = 2.72  on 1 df,   p=0.1
```

```
model <- coxph(Surv(months, vital) ~ factor(high_immune), tcga)
summary(model)
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ factor(high_immune), data = tcga)
##
##      n= 268, number of events= 149
##      (5 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_immune)1 0.1866      1.2051   0.2024 0.922      0.357
##
##              exp(coef) exp(-coef) lower .95 upper .95
```

```
## factor(high_immune)1      1.205      0.8298      0.8105      1.792
##
## Concordance= 0.496 (se = 0.018 )
## Likelihood ratio test= 0.82 on 1 df, p=0.4
## Wald test              = 0.85 on 1 df, p=0.4
## Score (logrank) test = 0.85 on 1 df, p=0.4
```

Analysis with covariates

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

Call:

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

##

```
##      n= 241, number of events= 136
```

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

##

	coef	exp(coef)	se(coef)	z	Pr(> z)
## 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

##

	exp(coef)	exp(-coef)	lower .95	upper .95
## factor(high_fibro)1	1.040	0.9611	0.6965	1.554
## factor(age)2	1.485	0.6733	0.1699	12.979
## factor(age)3	1.492	0.6705	0.1875	11.863
## factor(age)4	1.545	0.6472	0.2024	11.796
## factor(age)5	2.336	0.4281	0.3088	17.667
## factor(age)6	1.797	0.5565	0.2282	14.145
## factor(age)7	2.021	0.4948	0.2656	15.374
## factor(age)8	2.676	0.3737	0.3501	20.450
## factor(age)9	4.057	0.2465	0.5267	31.256
## factor(debulking)suboptimal	1.111	0.9003	0.7550	1.634
## factor(FewerStage)4	1.147	0.8721	0.6549	2.007

##

```
## Concordance= 0.63 (se = 0.026 )
```

```
## Likelihood ratio test= 13.39 on 11 df, p=0.3
```

```
## Wald test              = 14.05 on 11 df, p=0.2
```

```
## Score (logrank) test = 14.71 on 11 df, p=0.2
```

```
model <- coxph(Surv(months, vital) ~ (factor(high_immune) + factor(age)+
                                     factor(debulking)+ factor(FewerStage)), tcga)
summary(model)
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_immune) +
##       factor(age) + factor(debulking) + factor(FewerStage)), data = tcga)
##
## n= 241, number of events= 136
## (32 observations deleted due to missingness)
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_immune)1      0.38251   1.46597  0.22108 1.730  0.0836 .
## factor(age)2              0.38692   1.47244  1.09799 0.352  0.7245
## factor(age)3              0.39403   1.48295  1.04806 0.376  0.7069
## factor(age)4              0.37416   1.45378  1.02453 0.365  0.7150
## factor(age)5              0.82236   2.27587  1.02133 0.805  0.4207
## factor(age)6              0.48719   1.62774  1.04395 0.467  0.6407
## factor(age)7              0.67382   1.96172  1.02948 0.655  0.5128
## factor(age)8              1.02667   2.79174  1.03205 0.995  0.3198
## factor(age)9              1.44084   4.22423  1.03753 1.389  0.1649
## factor(debulking)suboptimal 0.07336   1.07612  0.19722 0.372  0.7099
## factor(FewerStage)4       0.18276   1.20053  0.28471 0.642  0.5209
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(high_immune)1      1.466      0.6821   0.9505   2.261
## factor(age)2              1.472      0.6791   0.1712  12.666
## factor(age)3              1.483      0.6743   0.1901  11.567
## factor(age)4              1.454      0.6879   0.1952  10.829
## factor(age)5              2.276      0.4394   0.3075  16.846
## factor(age)6              1.628      0.6143   0.2104  12.595
## factor(age)7              1.962      0.5098   0.2608  14.755
## factor(age)8              2.792      0.3582   0.3693  21.104
## factor(age)9              4.224      0.2367   0.5528  32.277
## factor(debulking)suboptimal 1.076      0.9293   0.7311   1.584
## factor(FewerStage)4       1.201      0.8330   0.6871   2.098
##
## Concordance= 0.615 (se = 0.026 )
## Likelihood ratio test= 16.19 on 11 df,  p=0.1
## Wald test              = 16.77 on 11 df,  p=0.1
## Score (logrank) test = 17.49 on 11 df,  p=0.09
```

Tothill

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

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

quantiles <- quantile(tothill$Immune)
```



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

Basic analysis

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ factor(high_fibro), data = tothill)
##
##      n= 145, number of events= 73
##      (2 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_fibro)1 0.4540    1.5746   0.2544 1.784   0.0743 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(high_fibro)1    1.575     0.6351    0.9563    2.593
##
## Concordance= 0.561 (se = 0.03 )
## Likelihood ratio test= 3  on 1 df,  p=0.08
## Wald test               = 3.18  on 1 df,  p=0.07
## Score (logrank) test = 3.24  on 1 df,  p=0.07
```

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ factor(high_immune), data = tothill)
##
##      n= 145, number of events= 73
##      (2 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_immune)1 -0.1292    0.8788   0.2791 -0.463   0.643
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(high_immune)1    0.8788     1.138    0.5086    1.519
##
## Concordance= 0.512 (se = 0.028 )
## Likelihood ratio test= 0.22  on 1 df,  p=0.6
## Wald test              = 0.21  on 1 df,  p=0.6
## Score (logrank) test = 0.21  on 1 df,  p=0.6
```

Analysis with covariates

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro) + factor(age) +
##      factor(debulking) + factor(FewerStage)), data = tothill)
##
##      n= 128, number of events= 65
##      (19 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## 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 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(high_fibro)1      1.7172   0.5823   0.98059   3.007
## factor(age)2      0.3833   2.6087   0.03631   4.047
## factor(age)3      1.1866   0.8427   0.26248   5.365
## factor(age)4      0.3273   3.0556   0.07540   1.420
## factor(age)5      0.5008   1.9967   0.13569   1.849
## factor(age)6      0.8301   1.2047   0.20408   3.377
## factor(age)7      0.2440   4.0978   0.04987   1.194
## factor(age)8      2.2225   0.4499   0.56745   8.705
## factor(age)9      0.7459   1.3406   0.16121   3.451
## factor(debulking)suboptimal 0.6677   1.4976   0.35691   1.249
## factor(FewerStage)2      1.4086   0.7099   0.18550  10.696
## factor(FewerStage)4      9.6446   0.1037   1.04043  89.404
##
## Concordance= 0.697 (se = 0.037 )
## Likelihood ratio test= 33.54 on 12 df,  p=8e-04
## Wald test              = 34.98 on 12 df,  p=5e-04
## Score (logrank) test = 40.17 on 12 df,  p=7e-05
```

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_immune) +
##      factor(age) + factor(debulking) + factor(FewerStage)), data = tothill)
##
##      n= 128, number of events= 65
##      (19 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
```

```
## factor(high_immune)1      -0.14485    0.86515    0.33319 -0.435    0.6638
## factor(age)2             -1.03809    0.35413    1.20164 -0.864    0.3876
## factor(age)3             -0.08096    0.92224    0.75528 -0.107    0.9146
## factor(age)4             -1.30483    0.27122    0.73811 -1.768    0.0771
## factor(age)5             -0.86535    0.42091    0.66019 -1.311    0.1899
## factor(age)6             -0.35568    0.70070    0.71082 -0.500    0.6168
## factor(age)7             -1.47839    0.22800    0.81763 -1.808    0.0706
## factor(age)8              0.65863    1.93215    0.68938  0.955    0.3394
## factor(age)9             -0.49586    0.60905    0.77226 -0.642    0.5208
## factor(debulking)suboptimal -0.36316    0.69548    0.32101 -1.131    0.2579
## factor(FewerStage)2       0.42877    1.53536    1.03226  0.415    0.6779
## factor(FewerStage)4       2.43811   11.45136    1.13264  2.153    0.0314 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(high_immune)1      0.8652    1.15586    0.45028    1.662
## factor(age)2              0.3541    2.82381    0.03360    3.733
## factor(age)3              0.9222    1.08432    0.20987    4.053
## factor(age)4              0.2712    3.68707    0.06383    1.152
## factor(age)5              0.4209    2.37583    0.11541    1.535
## factor(age)6              0.7007    1.42715    0.17397    2.822
## factor(age)7              0.2280    4.38588    0.04592    1.132
## factor(age)8              1.9322    0.51756    0.50031    7.462
## factor(age)9              0.6090    1.64191    0.13406    2.767
## factor(debulking)suboptimal 0.6955    1.43786    0.37071    1.305
## factor(FewerStage)2       1.5354    0.65131    0.20303   11.611
## factor(FewerStage)4      11.4514    0.08733    1.24381  105.429
##
## Concordance= 0.695 (se = 0.038 )
## Likelihood ratio test= 30.36 on 12 df,  p=0.002
## Wald test              = 31.53 on 12 df,  p=0.002
## Score (logrank) test = 36.6 on 12 df,  p=3e-04
```

All

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

quantiles <- quantile(composition$Immune)
q1 <- quantiles[2]
median <- quantiles[3]
q3 <- quantiles[4]
composition$high_immune <- ifelse(composition$Immune > q3, 1, 0)

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ factor(high_fibro), data = composition)
##
## n= 839, number of events= 444
## (16 observations deleted due to missingness)
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_fibro)1 0.2160    1.2411   0.1065 2.028  0.0426 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(high_fibro)1    1.241    0.8057    1.007    1.529
##
## Concordance= 0.523 (se = 0.012 )
## Likelihood ratio test= 3.98 on 1 df,  p=0.05
## Wald test               = 4.11 on 1 df,  p=0.04
## Score (logrank) test = 4.13 on 1 df,  p=0.04
```

```
model <- coxph(Surv(months, vital) ~ factor(high_immune), composition)
summary(model)
```

```
## Call:
## coxph(formula = Surv(months, vital) ~ factor(high_immune), data = composition)
##
## n= 839, number of events= 444
## (16 observations deleted due to missingness)
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_immune)1 -0.008399 0.991637 0.113868 -0.074  0.941
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(high_immune)1    0.9916    1.008    0.7933    1.24
##
## Concordance= 0.501 (se = 0.012 )
## Likelihood ratio test= 0.01 on 1 df,  p=0.9
## Wald test              = 0.01 on 1 df,  p=0.9
## Score (logrank) test = 0.01 on 1 df,  p=0.9
```

Analysis with covariates

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

```
## Call:
## coxph(formula = Surv(months, vital) ~ (factor(high_fibro) + factor(age) +
##   factor(debulking) + factor(FewerStage)), data = composition)
##
## n= 756, number of events= 408
## (99 observations deleted due to missingness)
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_fibro)1    0.11719   1.12434 0.11628 1.008  0.3135
```

```
## factor(age)2          0.04806    1.04924    0.46283 0.104    0.9173
## factor(age)3          0.03723    1.03793    0.42502 0.088    0.9302
## factor(age)4          0.14465    1.15563    0.40400 0.358    0.7203
## factor(age)5          0.39065    1.47795    0.39857 0.980    0.3270
## factor(age)6          0.41183    1.50958    0.40764 1.010    0.3124
## factor(age)7          0.24761    1.28096    0.40565 0.610    0.5416
## factor(age)8          0.72868    2.07234    0.40115 1.816    0.0693 .
## factor(age)9          0.96183    2.61649    0.40632 2.367    0.0179 *
## factor(debulking)suboptimal 0.13897    1.14909    0.10823 1.284    0.1991
## factor(FewerStage)2    0.62737    1.87267    0.71230 0.881    0.3784
## factor(FewerStage)4    0.93245    2.54072    0.72367 1.288    0.1976
## factor(FewerStage)9    0.79114    2.20592    1.23026 0.643    0.5202
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(high_fibro)1    1.124    0.8894    0.8952    1.412
## factor(age)2          1.049    0.9531    0.4236    2.599
## factor(age)3          1.038    0.9635    0.4512    2.387
## factor(age)4          1.156    0.8653    0.5235    2.551
## factor(age)5          1.478    0.6766    0.6767    3.228
## factor(age)6          1.510    0.6624    0.6790    3.356
## factor(age)7          1.281    0.7807    0.5784    2.837
## factor(age)8          2.072    0.4825    0.9441    4.549
## factor(age)9          2.616    0.3822    1.1799    5.802
## factor(debulking)suboptimal 1.149    0.8703    0.9295    1.421
## factor(FewerStage)2    1.873    0.5340    0.4636    7.564
## factor(FewerStage)4    2.541    0.3936    0.6151   10.494
## factor(FewerStage)9    2.206    0.4533    0.1979   24.592
```

```
##
```

```
## Concordance= 0.627 (se = 0.015 )
```

```
## Likelihood ratio test= 40.99 on 13 df, p=1e-04
```

```
## Wald test = 42.41 on 13 df, p=6e-05
```

```
## Score (logrank) test = 43.9 on 13 df, p=3e-05
```

```
model <- coxph(Surv(months, vital) ~ (factor(high_immune) + factor(age)+
                                     factor(debulking)+ factor(FewerStage)), composition)
summary(model)
```

```
## Call:
```

```
## coxph(formula = Surv(months, vital) ~ (factor(high_immune) +
##   factor(age) + factor(debulking) + factor(FewerStage)), data = composition)
```

```
##
```

```
## n= 756, number of events= 408
```

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

```
##
```

```
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(high_immune)1    0.176613  1.193169  0.122455  1.442  0.1492
## factor(age)2          -0.005217  0.994796  0.460855 -0.011  0.9910
## factor(age)3          -0.021051  0.979169  0.421499 -0.050  0.9602
## factor(age)4           0.079650  1.082908  0.401327  0.198  0.8427
## factor(age)5           0.311621  1.365637  0.396014  0.787  0.4313
## factor(age)6           0.322477  1.380543  0.405128  0.796  0.4260
## factor(age)7           0.199324  1.220578  0.405906  0.491  0.6234
```

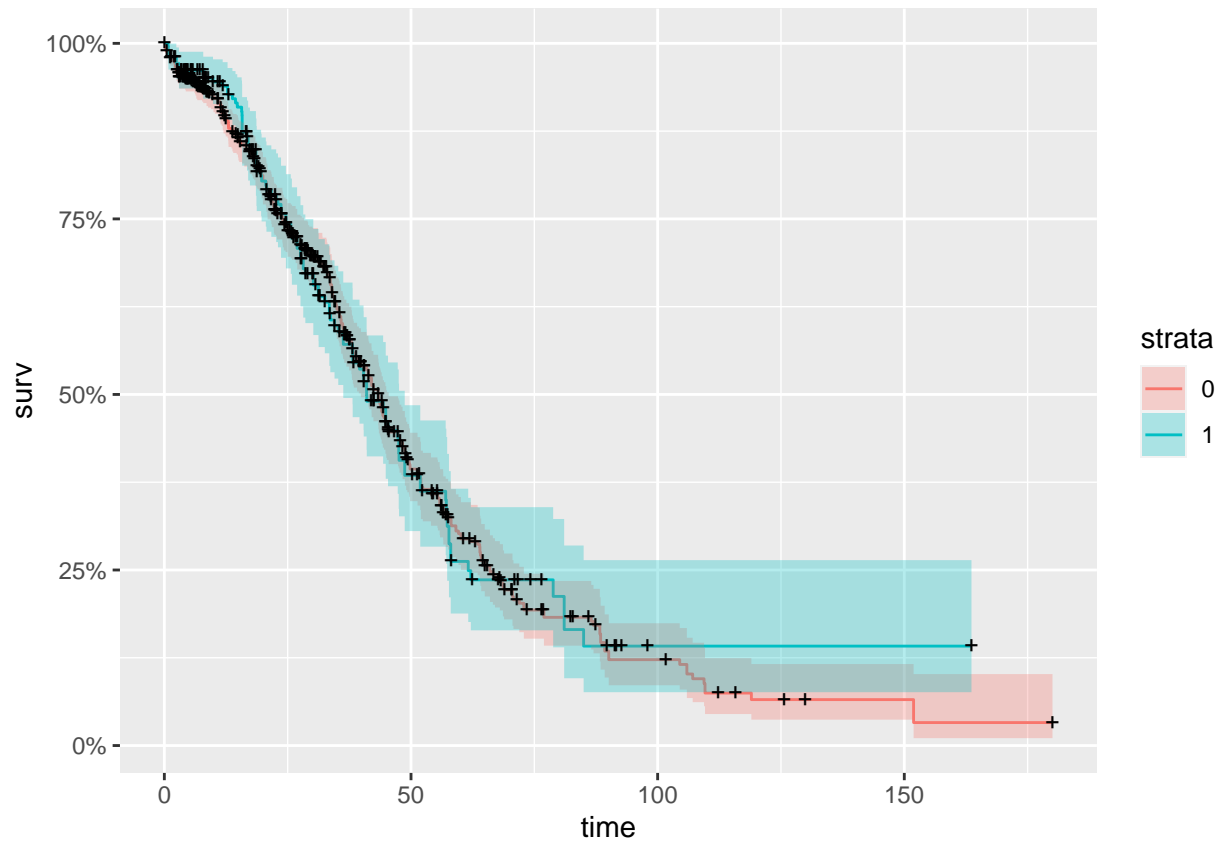
```

## factor(age)8          0.687930  1.989593  0.399831  1.721    0.0853 .
## factor(age)9          0.924395  2.520342  0.404454  2.286    0.0223 *
## factor(debulking)suboptimal 0.116683  1.123764  0.108942  1.071    0.2841
## factor(FewerStage)2    0.679369  1.972633  0.711810  0.954    0.3399
## factor(FewerStage)4    0.991124  2.694260  0.724151  1.369    0.1711
## factor(FewerStage)9    0.861870  2.367583  1.231417  0.700    0.4840
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(high_immune)1    1.1932    0.8381    0.9386    1.517
## factor(age)2           0.9948    1.0052    0.4031    2.455
## factor(age)3           0.9792    1.0213    0.4286    2.237
## factor(age)4           1.0829    0.9234    0.4932    2.378
## factor(age)5           1.3656    0.7323    0.6284    2.968
## factor(age)6           1.3805    0.7244    0.6240    3.054
## factor(age)7           1.2206    0.8193    0.5509    2.704
## factor(age)8           1.9896    0.5026    0.9087    4.356
## factor(age)9           2.5203    0.3968    1.1407    5.568
## factor(debulking)suboptimal 1.1238    0.8899    0.9077    1.391
## factor(FewerStage)2    1.9726    0.5069    0.4888    7.960
## factor(FewerStage)4    2.6943    0.3712    0.6517   11.139
## factor(FewerStage)9    2.3676    0.4224    0.2119   26.454
##
## Concordance= 0.626 (se = 0.015 )
## Likelihood ratio test= 42.01 on 13 df,  p=7e-05
## Wald test              = 43.54 on 13 df,  p=4e-05
## Score (logrank) test = 45.04 on 13 df,  p=2e-05

# Kaplan Meier plot for all datasets
km <- Surv(composition$months, composition$vital)
km_treatment<-survfit(km~high_immune,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). It's obviously not going to matter as much as stage or age, but it's still important to consider.