

Survival analysis of TCGA data based on RNA-Seq-derived immunophenotype

Refs:

- <https://stats.stackexchange.com/questions/29135/difference-between-survival-analysis-and-classification>
- <https://cran.r-project.org/web/packages/party/party.pdf> / ctree
- For a general description of the methodology see Hothorn, Hornik and Zeileis (2006) and Hothorn, Hornik, van de Wiel and Zeileis (2006). Introductions for novices can be found in Strobl et al. (2009) and at <http://github.com/christophM/overview-ctrees.git>.

Torsten Hothorn, Kurt Hornik, Mark A. van de Wiel and Achim Zeileis (2006). A Lego System for Conditional Inference. The American Statistician, 60(3), 257–263. Torsten Hothorn, Kurt Hornik and Achim Zeileis (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics, 15(3), 651–674. Preprint available from <http://statmath.wu-wien.ac.at/~zeileis/papers/Hothorn+Hornik+Zeileis-2006.pdf>

Load data

```
library(party)
```

```
## Loading required package: grid
## Loading required package: mvtnorm
## Loading required package: modeltools
## Loading required package: stats4
## Loading required package: strucchange
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
## Loading required package: sandwich
```

```
library(survival)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(memisc)
```

```
## Loading required package: lattice
```

```

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

##
## Attaching package: 'memisc'

## The following objects are masked from 'package:dplyr':
##
##     collect, query, recode, rename

## The following objects are masked from 'package:modeltools':
##
##     Lapply, relabel

## The following objects are masked from 'package:stats':
##
##     contrasts, contr.sum, contr.treatment

## The following object is masked from 'package:base':
##
##     as.array

library(reshape2)

df = read.table("Tree.txt", header = T, sep = "\t") %>% filter(!is.na(OS.uncorrected))

df$Stage2 = with(df, cases(
  Stage == "i" -> 1,
  Stage == "ii" -> 2,
  Stage == "iii" -> 3,
  Stage == "iv" -> 4,
  T -> NA
))

## Warning in cases(1 <- Stage == "i", 2 <- Stage == "ii", 3 <- Stage ==
## "iii", : conditions are not mutually exclusive

df$Tumor.site2 = with(df, cases(
  Tumor.site == "Distant Metastasis" -> "D.M.",
  Tumor.site == "Primary Tumor" -> "P.T.",
  Tumor.site == "Regional Cutaneous o" -> "R.C.",
  Tumor.site == "Regional Lymph Node" -> "R.L.N.",
  T -> NA
))

## Warning in cases("D.M." <- Tumor.site == "Distant Metastasis", "P.T." <-
## Tumor.site == : conditions are not mutually exclusive

df$Tumor.site2 = as.factor(df$Tumor.site2)

df$Stage3 = as.factor(with(df, ifelse(Stage == "" | Stage == "0", NA, as.character(Stage))))

ggplot(df %>% dplyr::select(IGH.clonality,
  IGH.coverage.by.MiXCR,

```

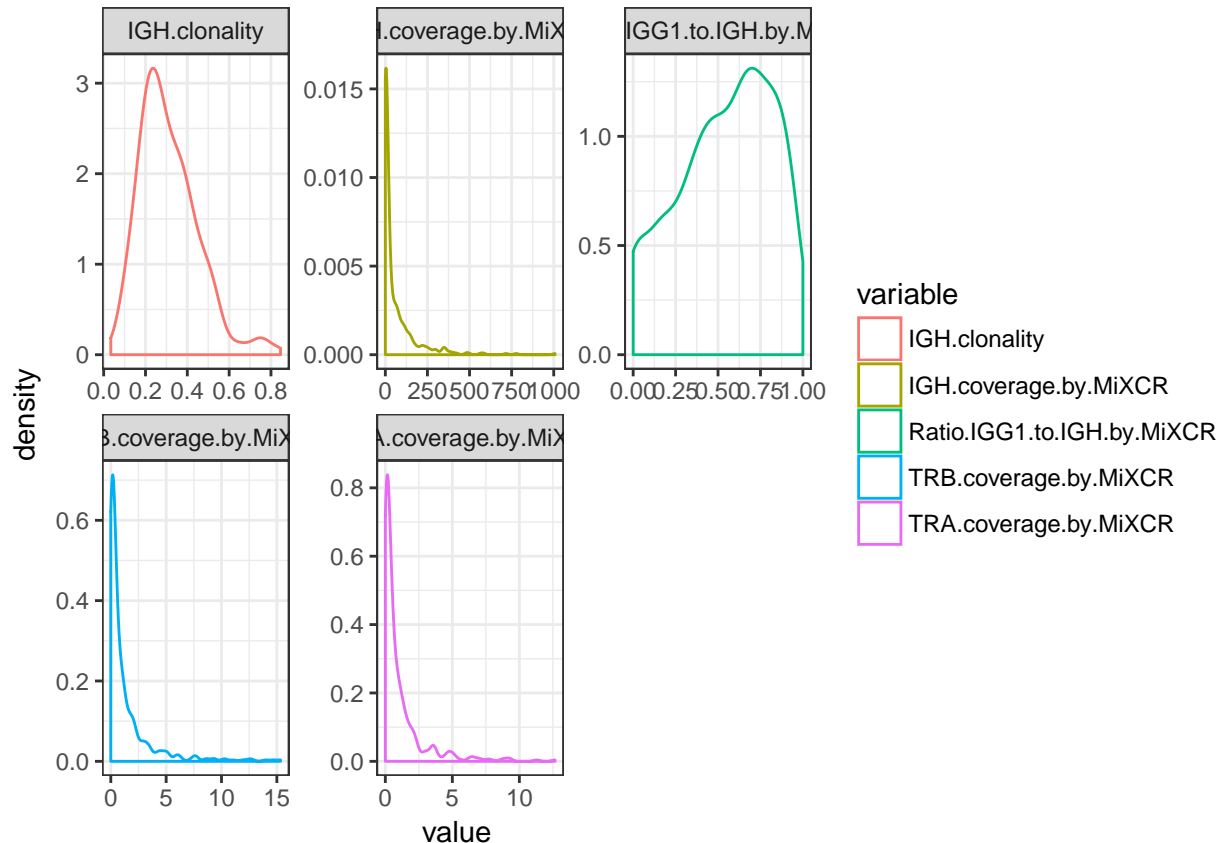
```

Ratio.IGG1.to.IGH.by.MiXCR,
TRB.coverage.by.MiXCR,
TRA.coverage.by.MiXCR) %>% melt,
aes(x=value, color=variable)) +
geom_density() +
facet_wrap(~variable, scales = "free") +
theme_bw()

```

No id variables; using all as measure variables

Warning: Removed 301 rows containing non-finite values (stat_density).



Decision tree learning parameters - we don't correct P-values for multiple testing; minimal size of bucket is 30 donors

```

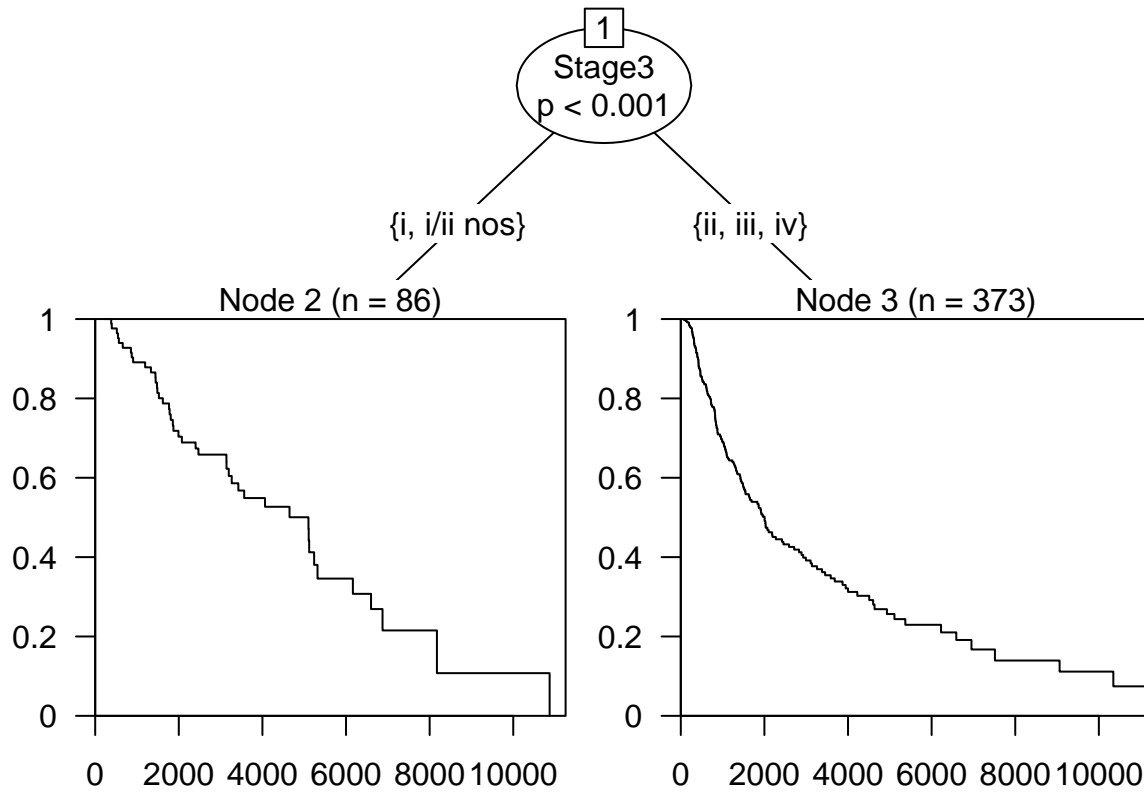
ctree_opts = ctree_control(testtype = "Univariate", #mincriterion = 0.8, #
                           minbucket = 45)

make_tree = function(formula) {
  dfct = ctree(formula, data = df, controls = ctree_opts)
  plot(dfct)
}

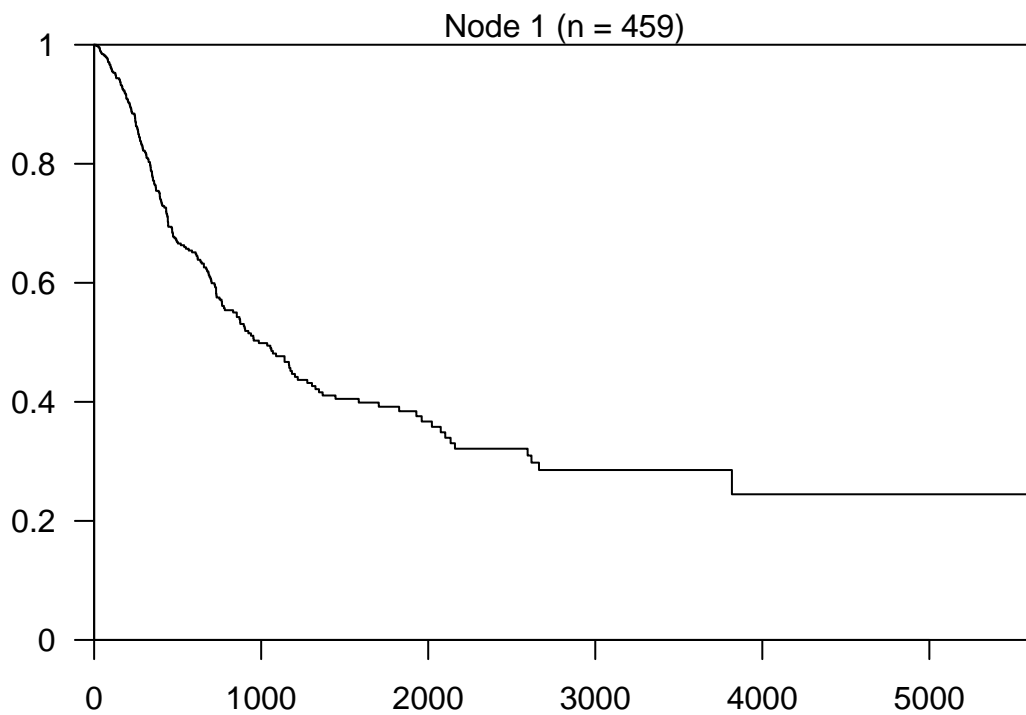
```

Stage correction for OS works well

```
make_tree(Surv(OS.uncorrected, Dead) ~ Stage3)
```

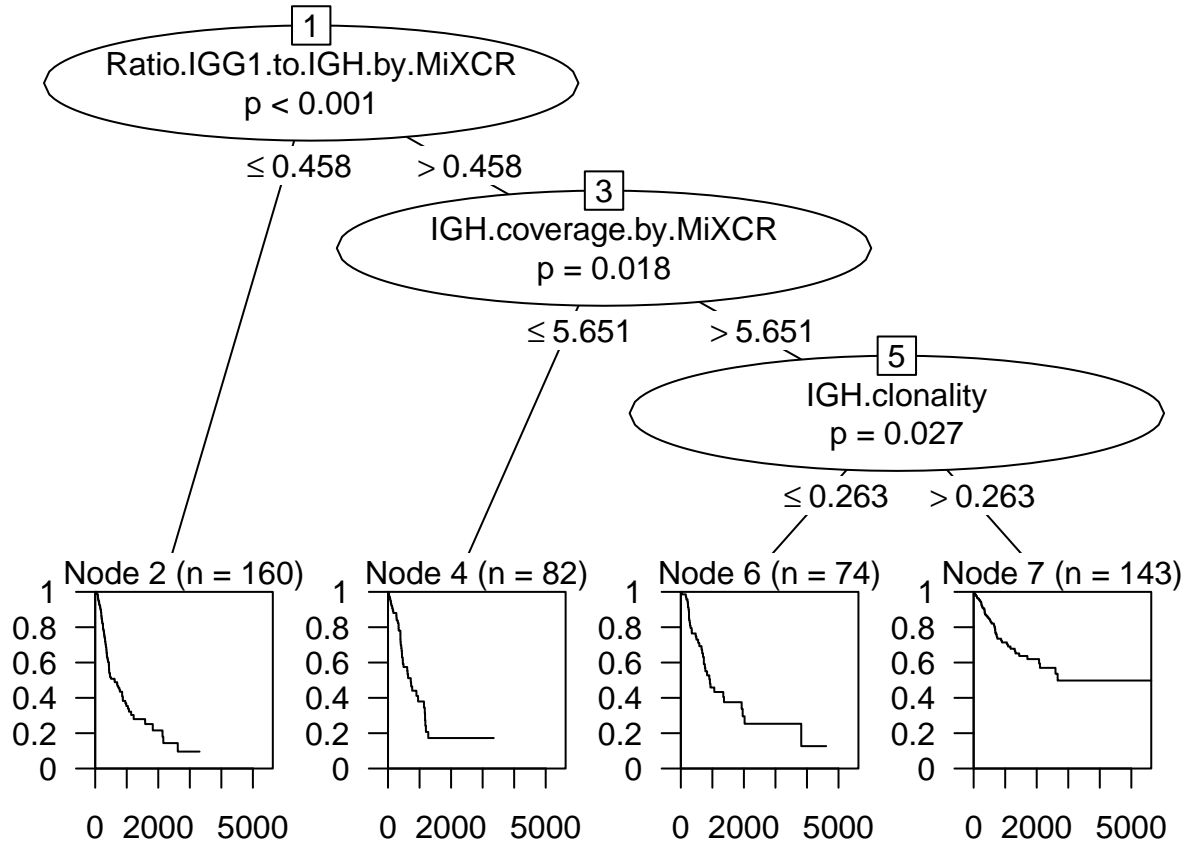


```
make_tree(Surv(OS.corrected, Dead) ~ Stage3)
```



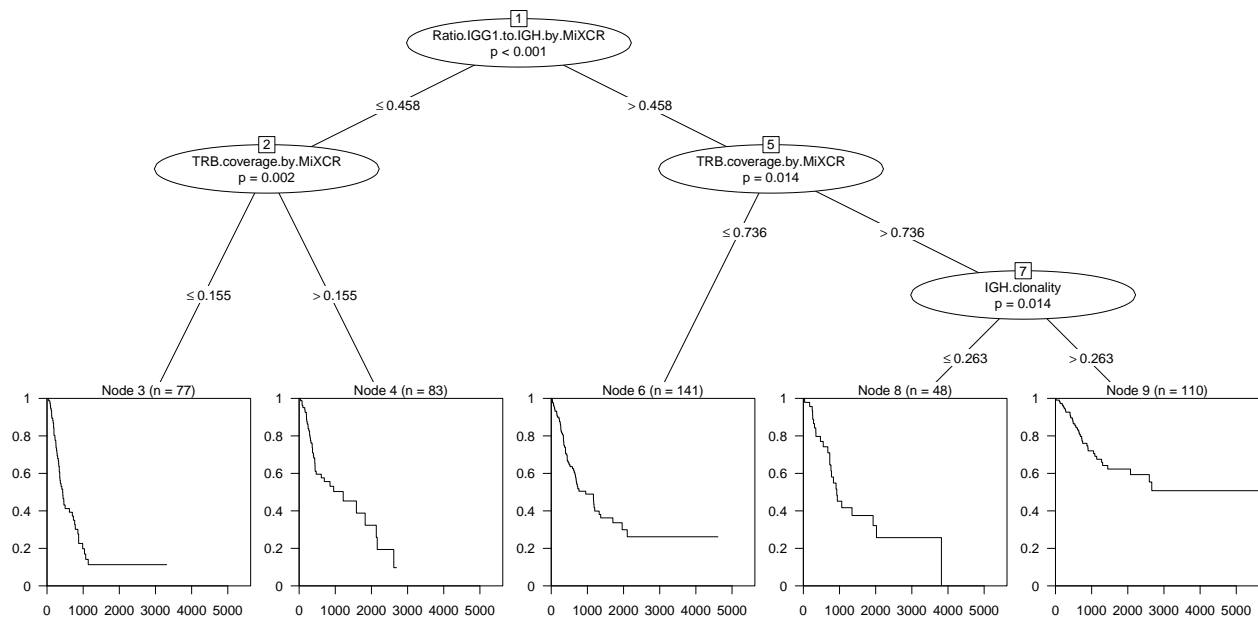
IG-based model plus tumor site (has no effect)

```
make_tree(Surv(OS.corrected, Dead) ~ Tumor.site2 +
  IGH.clonality +
  IGH.coverage.by.MiXCR +
  Ratio.IGG1.to.IGH.by.MiXCR)
```



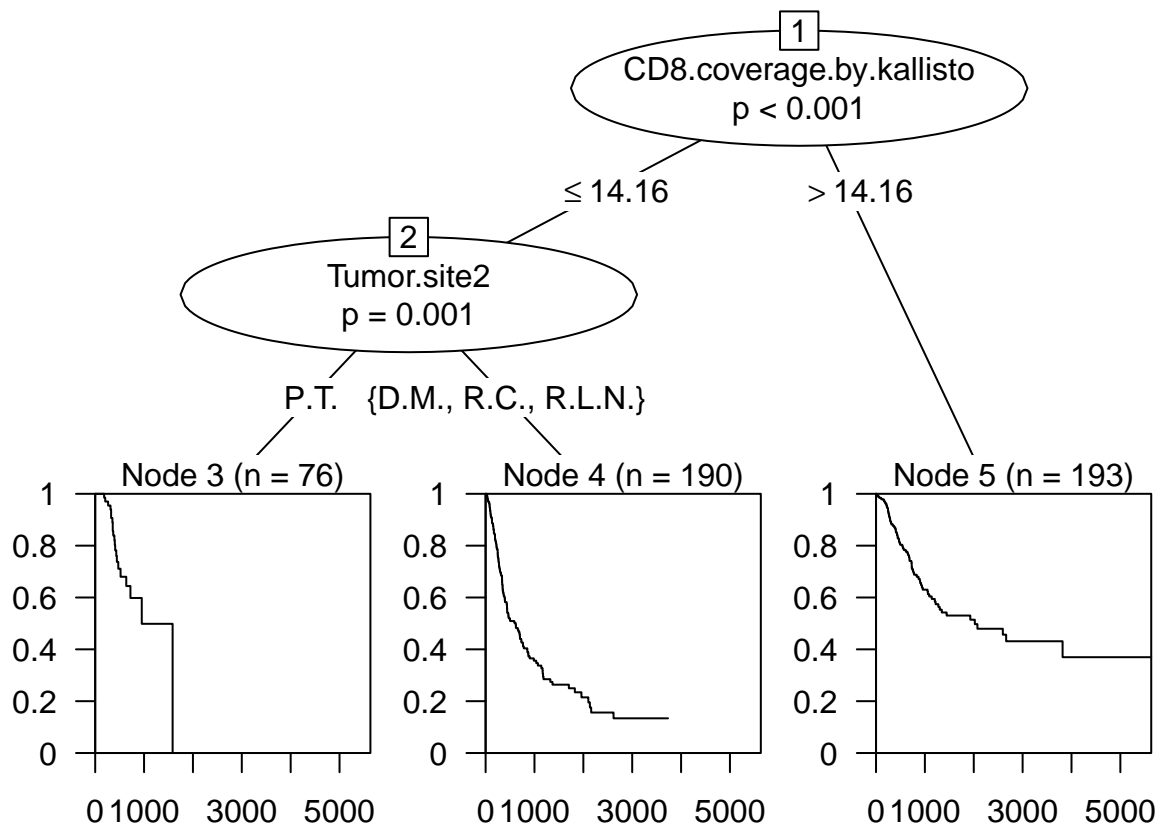
Adding TRB coverage and clonality

```
make_tree(Surv(OS.corrected, Dead) ~
  IGH.clonality +
  IGH.coverage.by.MiXCR +
  Ratio.IGG1.to.IGH.by.MiXCR +
  TRB.coverage.by.MiXCR +
  TRB.clonality)
```



Lets look who performs better among Clusters of Differentiation

```
make_tree(Surv(OS.corrected, Dead) ~
  Tumor.site2 +
  CD3.coverage.by.kallisto +
  CD8.coverage.by.kallisto +
  Ratio.CD138.to.CD20.by.kallisto +
  Ratio.CD138.to.CD45.by.kallisto +
  Ratio.CD319.to.CD45.by.kallisto)
```



Lets put everything together

```
make_tree(Surv(OS.corrected, Dead) ~
  Tumor.site2 +
  CD3.coverage.by.kallisto +
  CD8.coverage.by.kallisto +
  Ratio.CD138.to.CD20.by.kallisto +
  Ratio.CD138.to.CD45.by.kallisto +
  Ratio.CD319.to.CD45.by.kallisto +
  IGH.clonality +
  IGH.coverage.by.MiXCR +
  Ratio.IGG1.to.IGH.by.MiXCR +
  TRB.coverage.by.MiXCR +
  TRB.clonality)
```

