

Część 1

Agata Starzyk, Marta Czerwińska

Z dostępnych 33 różnych rodzajów raka, wybrałyśmy pięć z największą liczbą danych. Poniżej znajduje się podsumowanie danych.

typ raka	liczba obserwacji
KIPAN	757
GBMLGG	679
COADREAD	493
STES	483
KIRC	478
UCEC	440
OV	436
LGG	435
BRCA	410
LUAD	365
COAD	362
STAD	357
PRAD	352
BLCA	344
LUSC	328
GBM	244
SARC	226
THCA	224

typ raka	liczba obserwacji
KIRP	216
HNSC	212
SKCM	206
CESC	173
READ	131
ESCA	126
PAAD	123
TGCT	122
THYM	90
PCPG	82
KICH	63
LIHC	63
MESO	63
UCS	48
ACC	46
DLBC	33
CHOL	30
UVM	12

Wybranymi przez nas danymi są: GBMLGG, KIPAN, KIRC, STES, COADREAD, które odpowiednio oznaczają glioma wielopostaciowy mózgu niższej klasy, kohortowy rak nerki, rak żołądka i przełyku, rak chromofobowy nerki, gruczolakorak jelita grubego.

Aby zacząć część analityczną, musiałyśmy złączyć dane dotyczące ekspresji różnych białek z danymi dotyczącymi pacjentów z zaobserwowanymi odpowiednio rodzajami raka. Dane dotyczące pacjentów znajdują się w tabelkach z końcówką clinical, zaś z białkiem z końcówką RPPA.

Dodałyśmy zmienną time oznaczającą czas do ostatniej wizyty lub do śmierci oraz zmienną status oznaczającą status zajścia zdarzenia.

```
load("C:/Users/Marta/Desktop/Warsztaty/COADREAD.RPPA.rda")
load("C:/Users/Marta/Desktop/Warsztaty/COADREAD.clinical.rda")
COADREAD.RPPA$bcr_patient_barcode <- tolower(substr(COADREAD.RPPA$bcr_patient_barcode, 1, 12))

dane <- data.frame(COADREAD.clinical$patient.bcr_patient_barcode
, COADREAD.clinical$patient.days_to_death
, COADREAD.clinical$patient.days_to_last_followup
, time = ifelse(is.na(COADREAD.clinical$patient.days_to_death)
, COADREAD.clinical$patient.days_to_last_followup
, COADREAD.clinical$patient.days_to_death)
, status = ifelse(is.na(COADREAD.clinical$patient.days_to_death), 'alive', 'dead'))
```

```
dane$time <- as.numeric(as.character(dane$time))
allCOADREAD <- merge(dane, COADREAD.RPPA
                      , by.x = "COADREAD.clinical.patient.bcr_patient_barcode"
                      , by.y = "bcr_patient_barcode")
colnames(allCOADREAD)[which(names(allCOADREAD) ==
                             "COADREAD.clinical.patient.bcr_patient_barcode")] <- "barcode"
colnames(allCOADREAD)[which(names(allCOADREAD) ==
                             "COADREAD.clinical.patient.days_to_death")] <- "patient.days_to_death"
```

Powyższą czynność powtórzyliśmy dla każdego rodzaju raka.

Następnie dla każdego zbioru danych policzyliśmy p wartość każdego białka. Dane obliczenia znajdują się poniżej:

```
n <- dim(allCOADREAD)[2]
m <- dim(allCOADREAD)[1]
pwartosc <- numeric(n-6)
for (i in 6:n)
{
  val <- colnames(allCOADREAD)[i]
  allCOADREAD[,val] <- as.numeric(as.character(allCOADREAD[,val]))
  allCOADREAD[, "patient.days_to_death"] <- as.numeric(as.character(allCOADREAD$patient.days_to_death))

  if (sum(is.na(allCOADREAD[,val])) > m/2){
    pwartosc[i-6] <- NA
    next
  }
  else{
    med <- median(allCOADREAD[,val], na.rm=TRUE)
    if(med != all(allCOADREAD$val) || is.na(allCOADREAD$val)){
      a <- survdiff(Surv(time,status == "dead")~ (allCOADREAD[,val]>med), data=allCOADREAD)
      pwartosc[i-6] <- 1-pchisq(a$chisq,1)
    }
    else
    {
      pwartosc[i-6] <- 1
    }
  }
}
}
```

Z powyższych obliczonych p-value wybieramy pięć najmniejszych, innymi słowy wybieramy pięć najbardziej istotnych białek dla danego typu raka. Wybrane białka to Akt, B-Raf, CD31, Chk1, Myosin-IIa. Poniżej zamieszczamy wykresy krzywych przeżycia dla wybranych białek.

```
a <- which(pwartosc %in% sort(pwartosc)[1:5])
nazwy <- as.vector(colnames(allCOADREAD)[6:n])
wybrane <- nazwy[a]

i=1
for (i in 1:5)
{
  val <- wybrane[i]
```

```

allCOADREAD[,val] <- as.numeric(as.character(allCOADREAD[,val]))
allCOADREAD[,"patient.days_to_death"] <- as.numeric(as.character(allCOADREAD$patient.days_to_death))
med <- median(allCOADREAD[,val],na.rm=TRUE)
x <- survfit(Surv(time,status == "dead")~ (allCOADREAD[,val]>med), data=allCOADREAD)
library(ggplot2)
library(survMisc)
w <- autoplot(x,legend=FALSE, title = paste("Krzywe przeżycia dla białka",val))
ggsave(file = paste0("Białko.", val, ".png"), plot = w$plot)
print(w$plot)
}

```

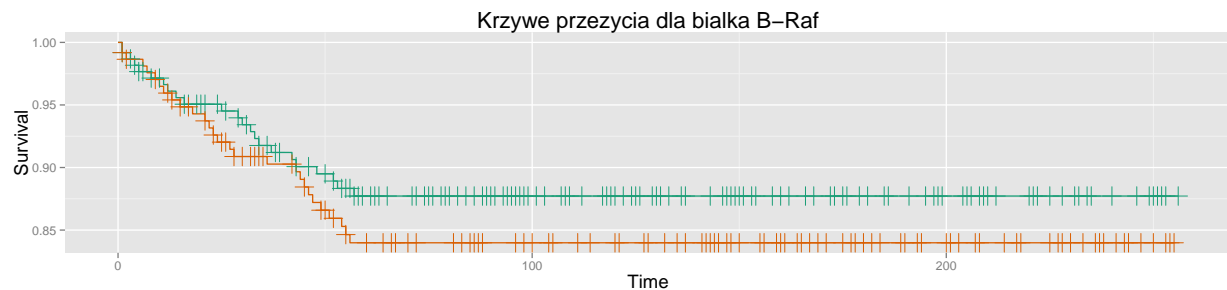
```

##
## Attaching package: 'survMisc'
##
## The following object is masked from 'package:ggplot2':
##
##   autoplot
##
## The following objects are masked from 'package:stats':
##
##   AIC, BIC, median, quantile
##
## Saving 14 x 3.5 in image
## Saving 14 x 3.5 in image

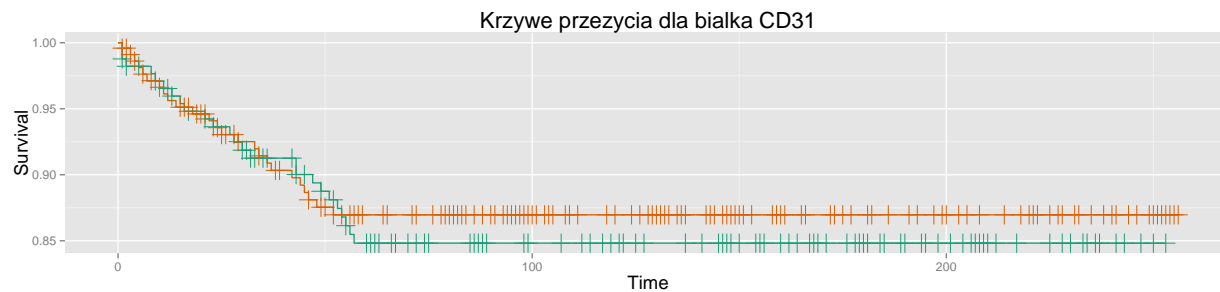
```



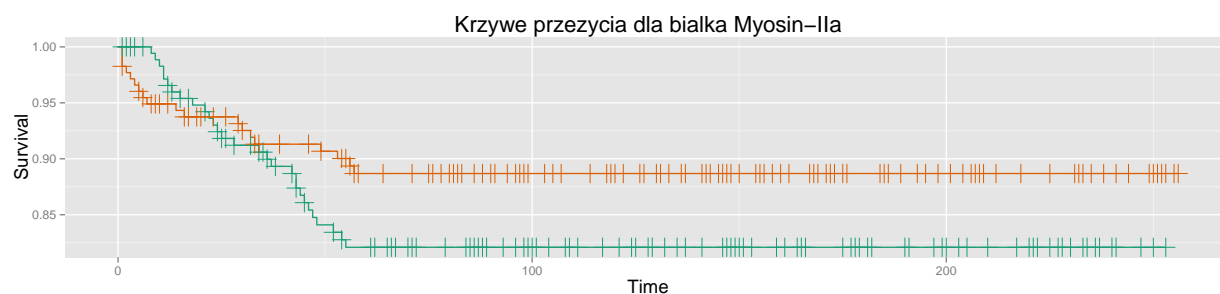
```
## Saving 14 x 3.5 in image
```



```
## Saving 14 x 3.5 in image
```



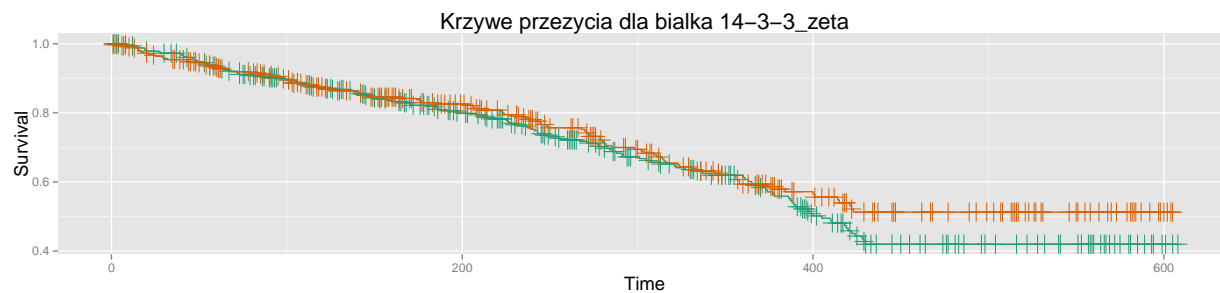
Saving 14 x 3.5 in image



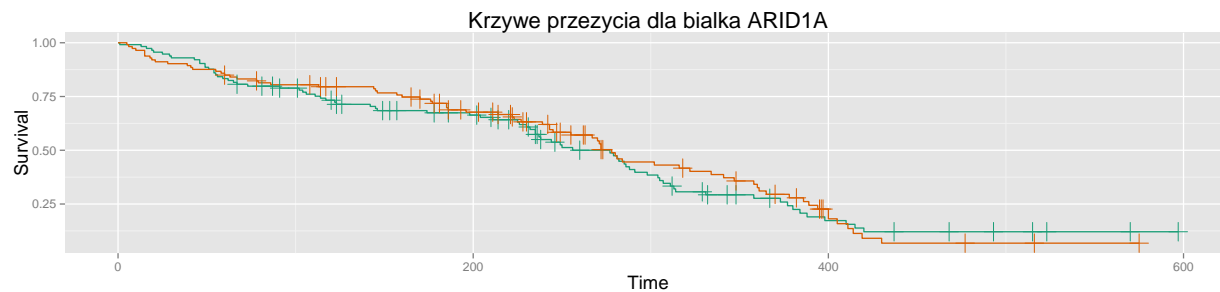
Analogicznie przeprowadziliśmy analizę dla glejaka wielopostaciowego mózgu niższej klasy. Wybrałyśmy następujące rodzaje białek: 14-3-3_zeta, ARID1A, Bap1-c-4, SCD, TSC1. Poniżej zamieszczamy wykresy krzywych przeżycia dla wybranych białek.

Saving 14 x 3.5 in image

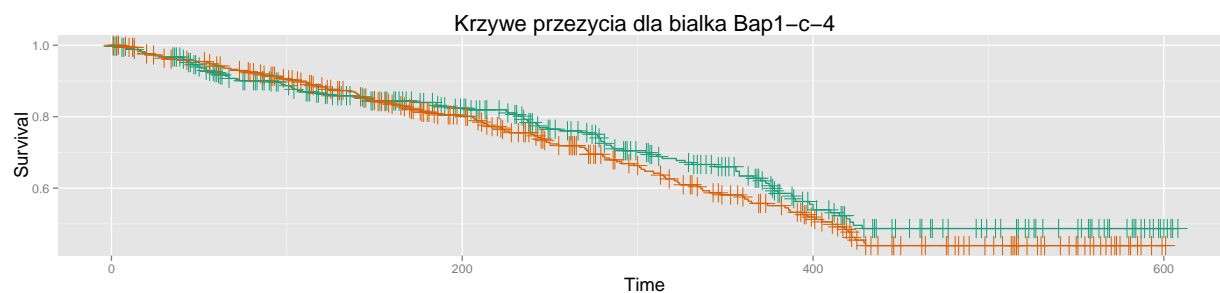
Saving 14 x 3.5 in image



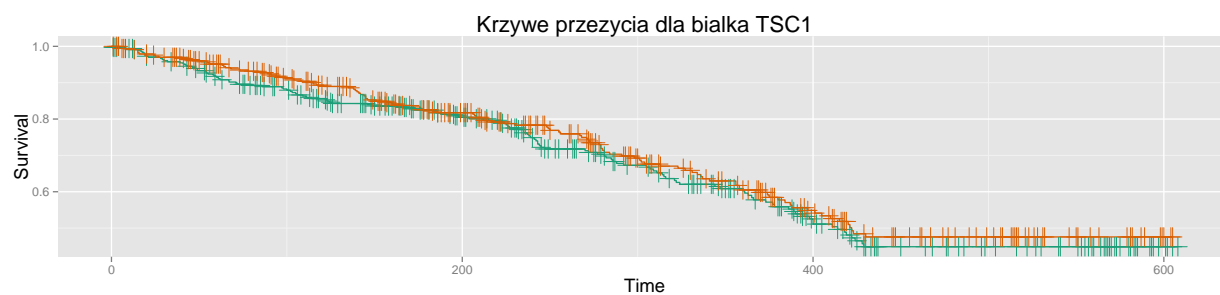
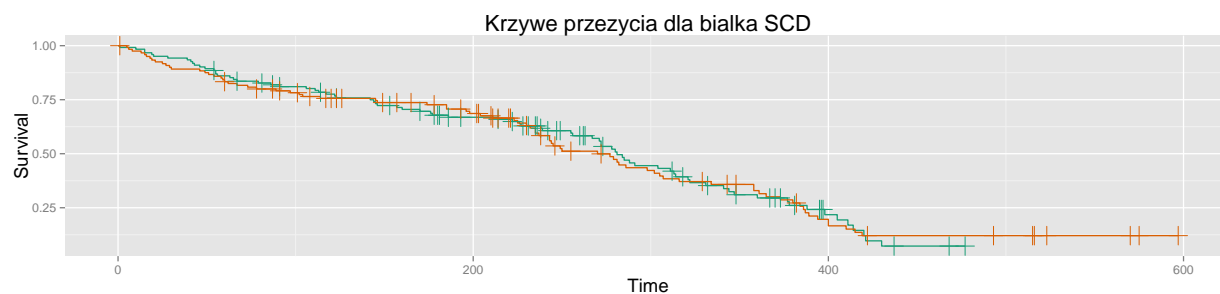
Saving 14 x 3.5 in image



Saving 14 x 3.5 in image



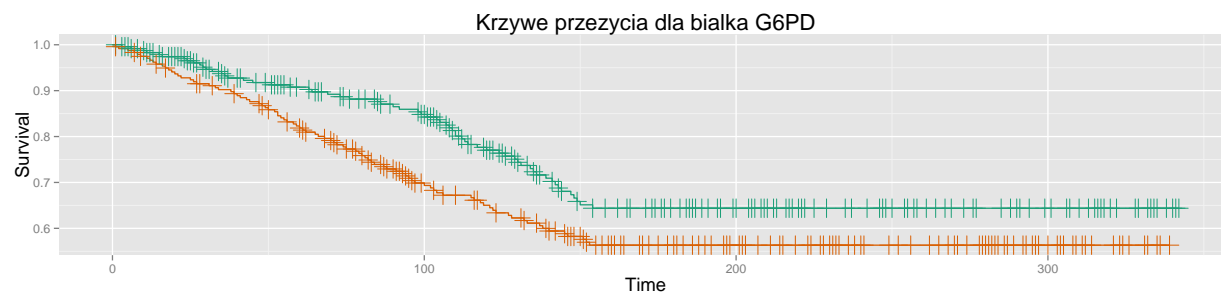
Saving 14 x 3.5 in image



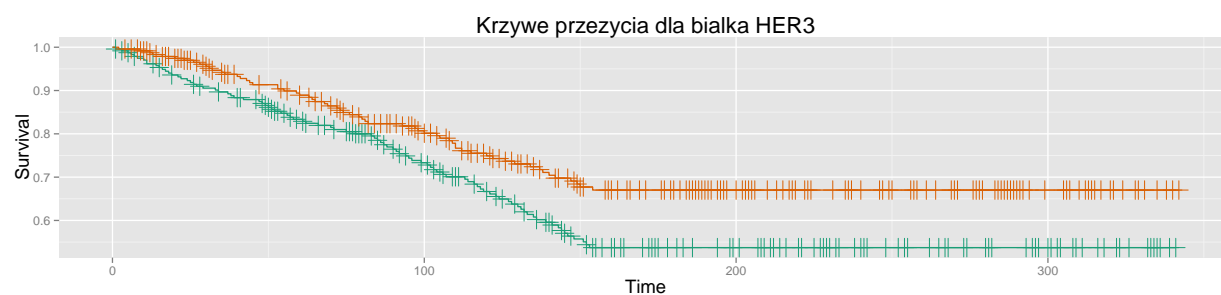
Białka, które wybrałyśmy dla raka żołądka i przelyku to: G6PD, HER3, MEK1-pS217-S221, Oxphos-complex-V_subunitb, cIAP. Krzywe KM wyglądają następująco:

Saving 14 x 3.5 in image

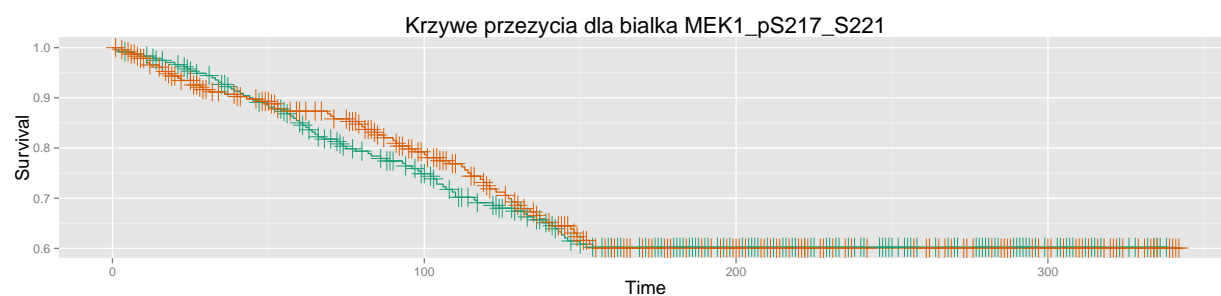
Saving 14 x 3.5 in image



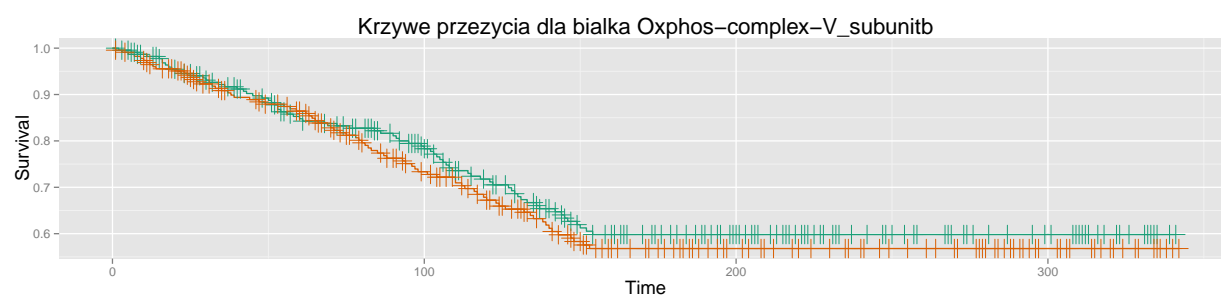
Saving 14 x 3.5 in image

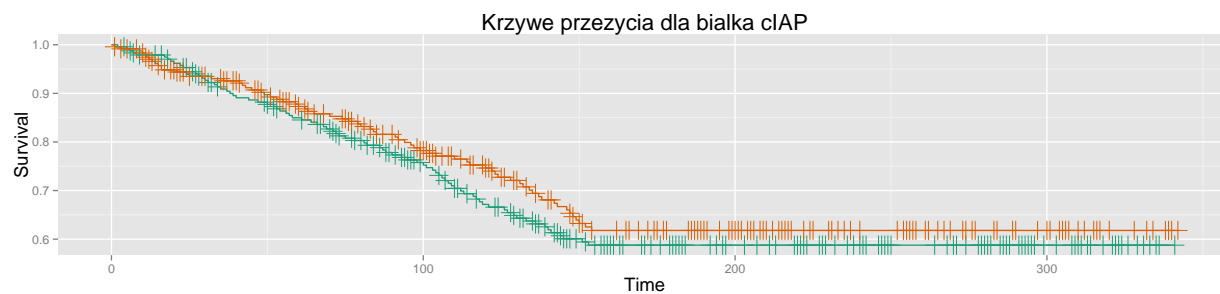


Saving 14 x 3.5 in image



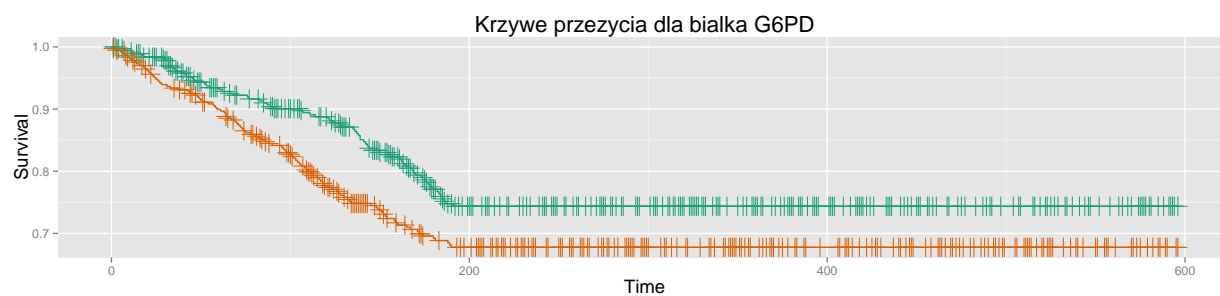
Saving 14 x 3.5 in image



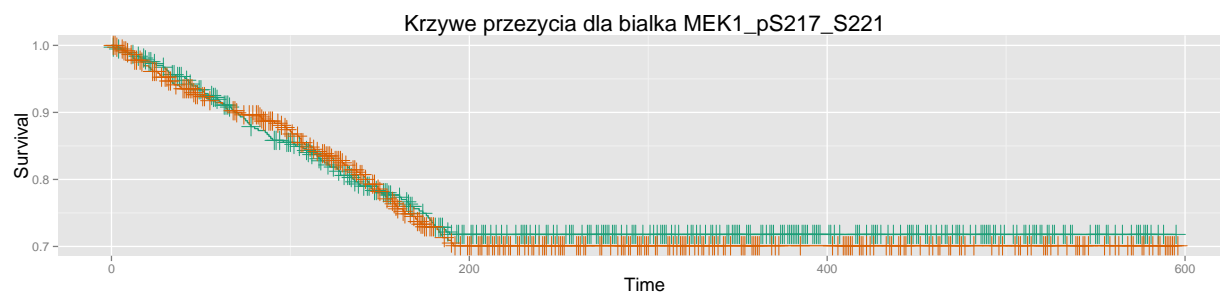


Kohortowy rak nerki: G6PD, MEK1-pS217-S221, TAZ, Tuberin, cIAP

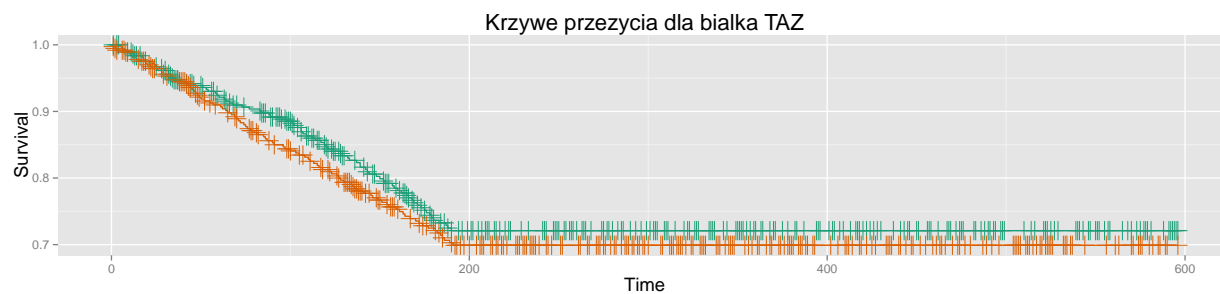
Saving 14 x 3.5 in image
Saving 14 x 3.5 in image



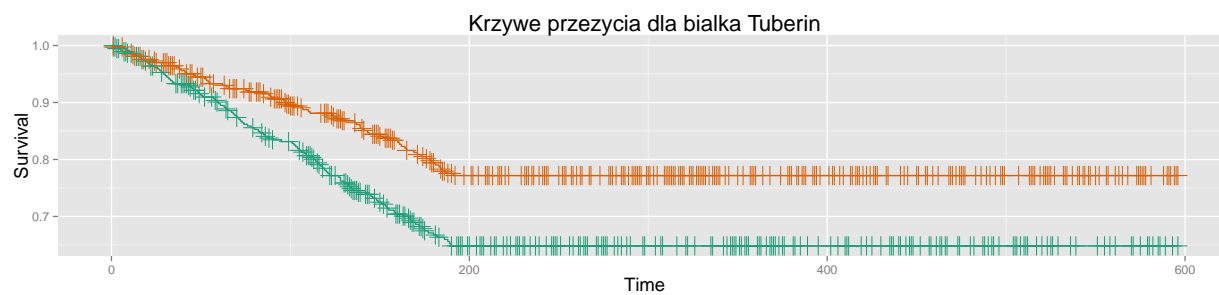
Saving 14 x 3.5 in image



Saving 14 x 3.5 in image



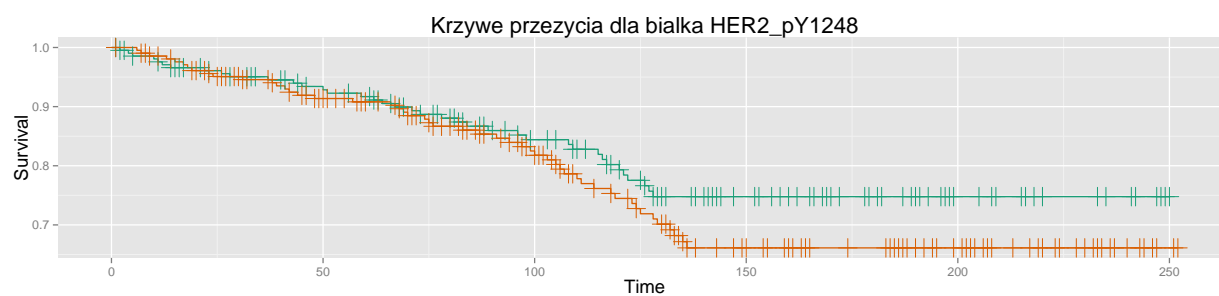
Saving 14 x 3.5 in image



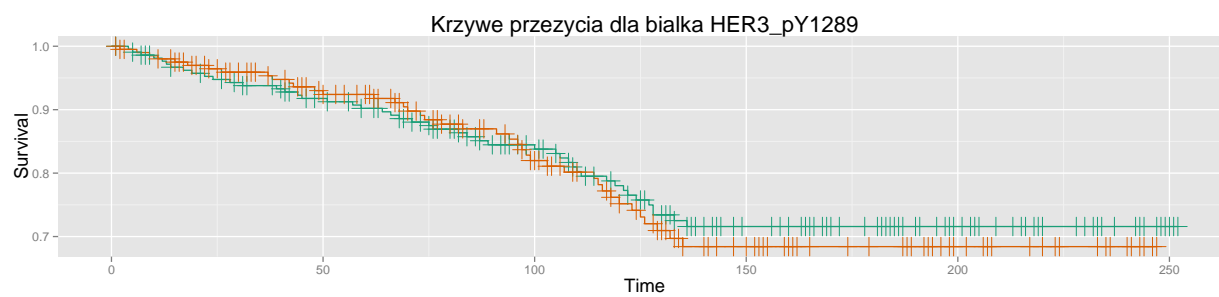
Dla raka chromofobowego nerki: bialka: HER2-pY1248, HER2-pY1289, PDK1, Rsc-pY416, p21, krzywe KM:

Saving 14 x 3.5 in image

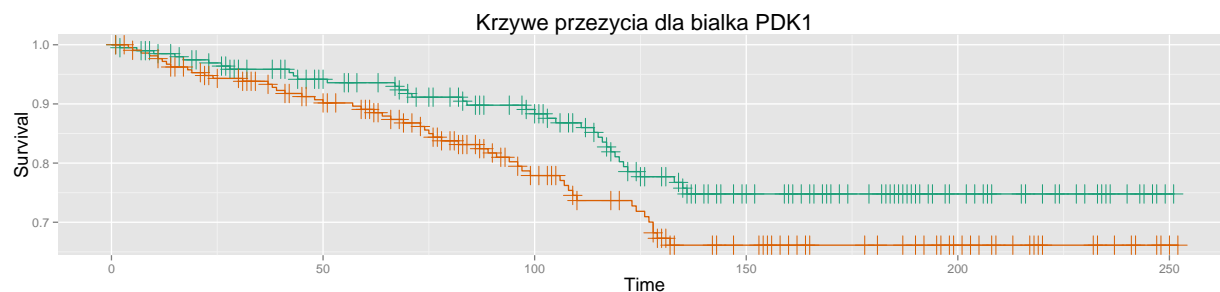
Saving 14 x 3.5 in image



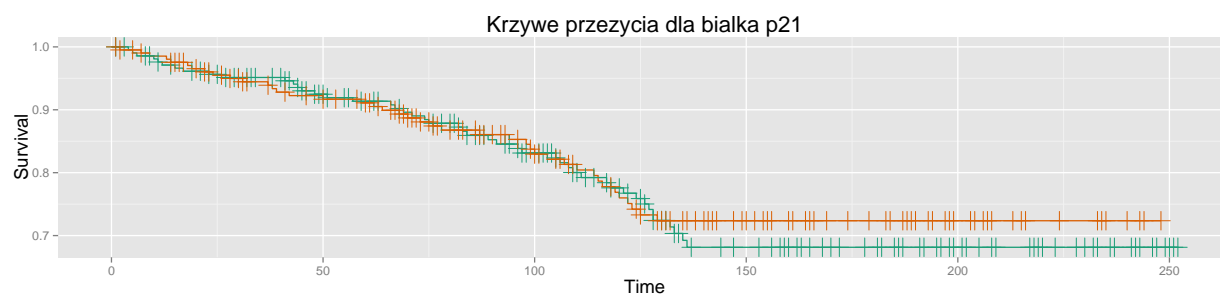
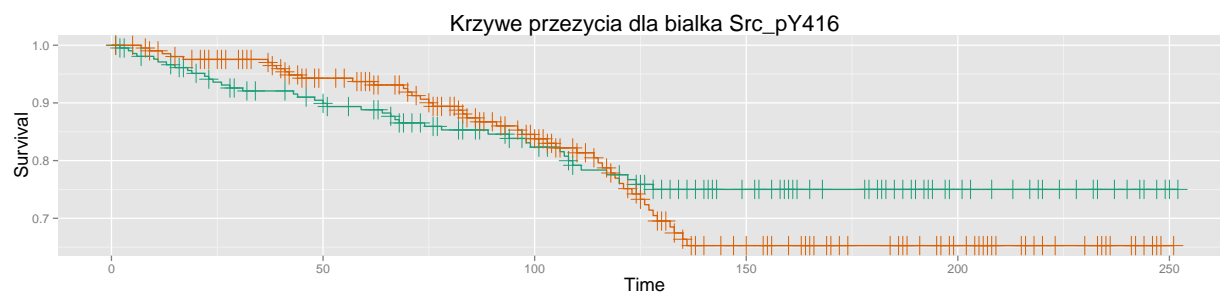
Saving 14 x 3.5 in image



Saving 14 x 3.5 in image



Saving 14 x 3.5 in image



Z przeprowadzonej przez nas analizy wynika, że najbardziej istotne i powtarzające się dla różnych typów białka to: 1. Myosin-IIa dla glejaka wielopostaciowego mózgu niższej klasy i gruczolakoraka jelita grubego; 2. G6PD dla kohortowego rak nerki oraz dla raka żołądka i przełyku; 3. MEK1_pS217_S221 dla kohortowego rak nerki oraz dla raka żołądka i przełyku; 4. cIAP dla kohortowego rak nerki oraz dla raka żołądka i przełyku.