TCGA BRCA CNV Analysis

Asad

5/5/2023

```
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(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.4.2     v purrr   1.0.1
## v tibble  3.2.1     v stringr 1.5.0
## v tidyr   1.3.0     v forcats 0.5.1
## v readr   2.1.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(rcartocolor)
library(colorspace)
library(ggpubr)
library(RColorBrewer)
library(wesanderson)
library(survival)
library(survminer)
```

```
##
## Attaching package: 'survminer'
```

```
## The following object is masked from 'package:survival':
##
##     myeloma

library(pROC)


## Type 'citation("pROC")' for a citation.


##
## Attaching package: 'pROC'


## The following object is masked from 'package:colorspace':
##
##     coords


## The following objects are masked from 'package:stats':
##
##     cov, smooth, var

library(scales)


##
## Attaching package: 'scales'


## The following object is masked from 'package:purrr':
##
##     discard


## The following object is masked from 'package:readr':
##
##     col_factor

library(patchwork)

#Set working directory and read data
setwd('D:/CancerData/TCGA BRCA cBioPortal/BRCA/Mutation')
cnv<- read.table('data_cna.txt', header = T, sep = '\t')

#Read the RNAseq cancer data that we prepared before
genes<-read.csv('log2_cancer.csv')

#Let's design the experiment for FOXM1 and MCM2
cnv_refined<- t(filter(cnv, cnv$Hugo_Symbol=='FOXM1' | cnv$Hugo_Symbol=='MCM2'))
#write.csv(cnv_refined, 'cnv_refined.csv')
cnv_refined<-read.csv('cnv_refined.csv')

#Merge expression and cnv datasets
Merged_cnv<- left_join(genes, cnv_refined, by='PATIENT_ID')

#Remove blanks and NA values
Merged_cnv[Merged_cnv=='']<-NA
```

```r
Merged_cnv<- na.omit(Merged_cnv)


#Let's make new columns according to alteration type
#[Remeber, -2= Deep Deletion, -1= Shallow Deletion, 0= No Alteration, 1=Gain, 2=Large Amplification]
Merged_cnv$FOXM1_Sum<- ifelse(Merged_cnv$FOXM1.CNV=='0','Wild Type', 'Altered' )
Merged_cnv$MCM2_Sum<- ifelse(Merged_cnv$MCM2.CNV=='0','Wild Type', 'Altered')


#Rename rows based on alteration type
Merged<- Merged_cnv #Make new object
#For Deletion
val_repl<- c(-2,-1)
col_repl<- c("FOXM1.CNV", "MCM2.CNV")

Merged[col_repl] <- sapply(Merged[col_repl],  # Replace values in certain columns
                           function(x) replace(x, x %in% val_repl, "Deletion"))
#For Amplification
val_repl<- c(2,1)
col_repl<- c("FOXM1.CNV", "MCM2.CNV")

Merged[col_repl] <- sapply(Merged[col_repl],  # Replace values in certain columns
                          function(x) replace(x, x %in% val_repl, "Amplification"))

Merged[Merged=='0']<- 'Wild Type'

#Save
#write.csv(Merged, 'cnv_final.csv')
Merged <- read.csv('cnv_final.csv')


#FOXM1 CNV Percentage Calculation
FOXM1_CNV <- Merged %>%
  group_by(FOXM1.CNV) %>% # Variable to be transformed
  count() %>%
  ungroup() %>%
  mutate(perc = `n` / sum(`n`)) %>%
  arrange(perc) %>%
  mutate(labels = scales::percent(perc))

#MCM2 CNV Percentage Calculation
MCM2_CNV <- Merged %>%
  group_by(MCM2.CNV) %>% # Variable to be transformed
  count() %>%
  ungroup() %>%
  mutate(perc = `n` / sum(`n`)) %>%
  arrange(perc) %>%
  mutate(labels = scales::percent(perc))

#Plotting
FOXM1_Plot<- ggplot(FOXM1_CNV, aes(x = "", y = perc, fill = FOXM1.CNV)) +
                      geom_col(color = "black") + geom_label(aes(label =
                        labels), color = c(1, "black", "black"),
```

```
                                     position = position_stack(vjust = 0.5),
                                  show.legend = F) + guides(fill = guide_legend(title = "FOXM1 CNV Summary"))
                               scale_fill_manual(values=brewer.pal(n = 3, name = "Set2"))+
                                  coord_polar(theta = "y") +
                                theme_void()

FOXM1_Plot
```
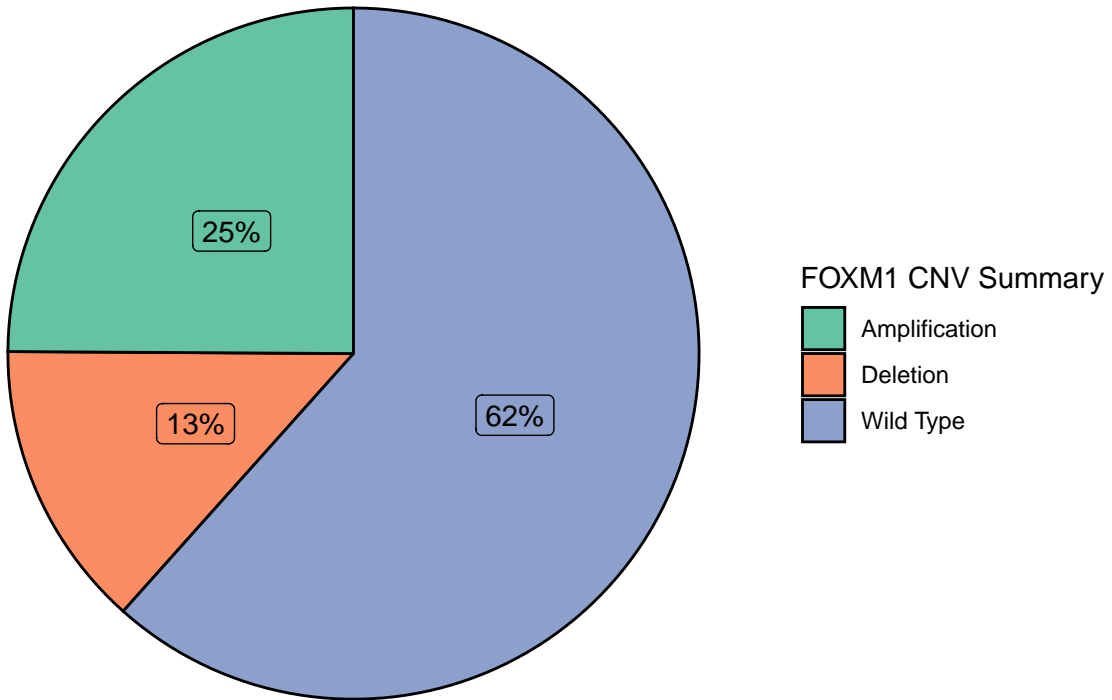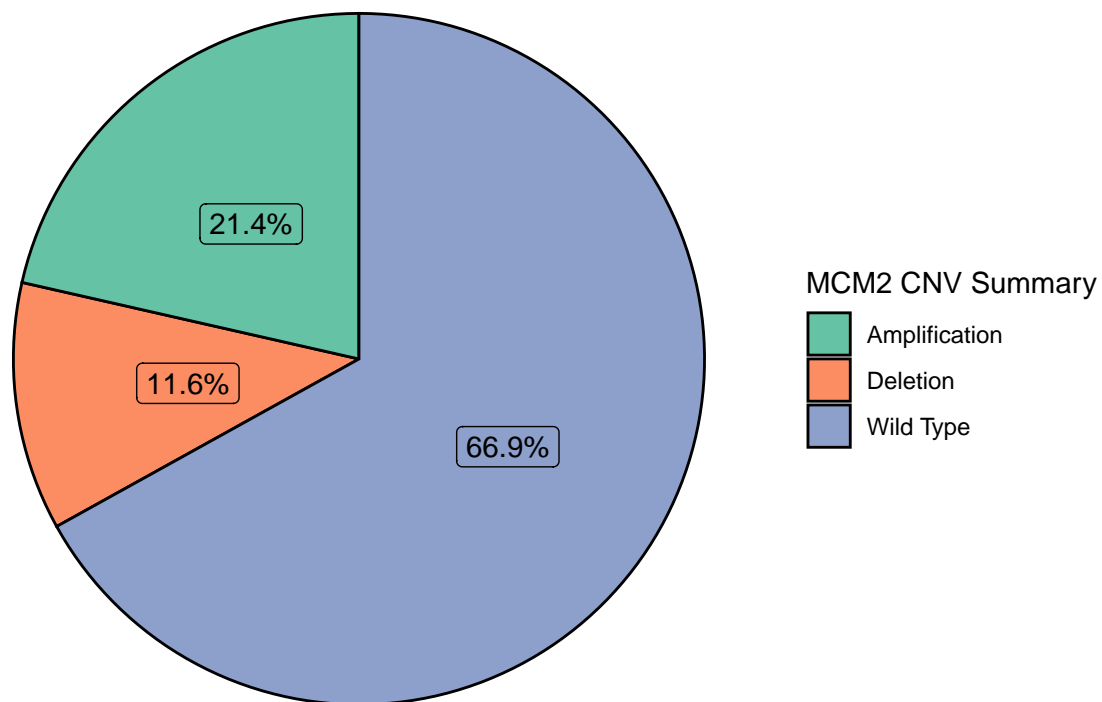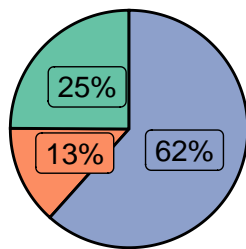


```
MCM2_Plot<- ggplot(MCM2_CNV, aes(x = "", y = perc, fill = MCM2.CNV)) +
        geom_col(color = "black") +
          geom_label(aes(label = labels), color = c(1, "black", "black"),
            position = position_stack(vjust = 0.5),
            show.legend = F) +
          guides(fill = guide_legend(title = "MCM2 CNV Summary")) +
            scale_fill_manual(values=brewer.pal(n = 3, name = "Set2"))+
        coord_polar(theta = "y") +
          theme_void()
MCM2_Plot
```
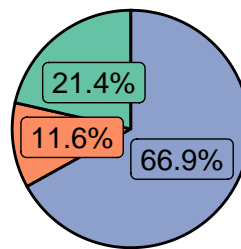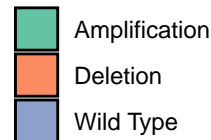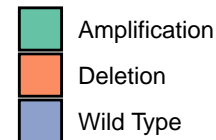
MCM2 CNV Summary
- Amplification
- Deletion
- Wild Type

21.4%

11.6%

66.9%

FOXM1_Plot+MCM2_Plot

FOXM1 CNV Summary
- Amplification
- Deletion
- Wild Type

25% 13% 62%

MCM2 CNV Summary
- Amplification
- Deletion
- Wild Type

21.4% 11.6% 66.9%

```r
#Examining impact on gene expression
my_pal <- rcartocolor::carto_pal(n = 8, name = "Bold")[c(1, 3, 7, 2)]

#FOXM1 Expression
FOXM1_Expr <- ggplot(Merged, aes(x =factor(FOXM1.CNV, levels = c('Wild Type', 'Deletion', 'Amplification
                      , y = FOXM1, color = FOXM1.CNV, fill = FOXM1.CNV)) +
                scale_y_continuous() + labs(x="CNV Type", y='FOXM1 Expression in log2(RSEM+1)')+
                 scale_color_manual(values = my_pal, guide = "none") + scale_fill_manual(values = 
                 guide = "none")


FOXM1_Expr<- FOXM1_Expr + theme(axis.title.x =  element_text(size = 14), axis.title.y =
           (element_text(size = 12)), axis.text = element_text(size = 12), panel.background =
             element_rect(fill = 'NA', color = 'white'),panel.border = element_rect(colour =
             'black', size = 1, fill = 'NA'))
```

```
## Warning: The 'size' argument of 'element_rect()' is deprecated as of ggplot2 3.4.0.
## i Please use the 'linewidth' argument instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```r
FOXM1_Expr<-FOXM1_Expr +
  ggbeeswarm::geom_quasirandom(
    size = 2, width = .33, alpha = .3
```

```
  ) +
  stat_summary(
    fun = median, geom = "point",
    shape = 95, size = 50
  ) +
  ggbeeswarm::geom_quasirandom(
    size = 2, width = .33, shape = 1, color = "black", stroke = .8
  )
FOXM1_Expr
```
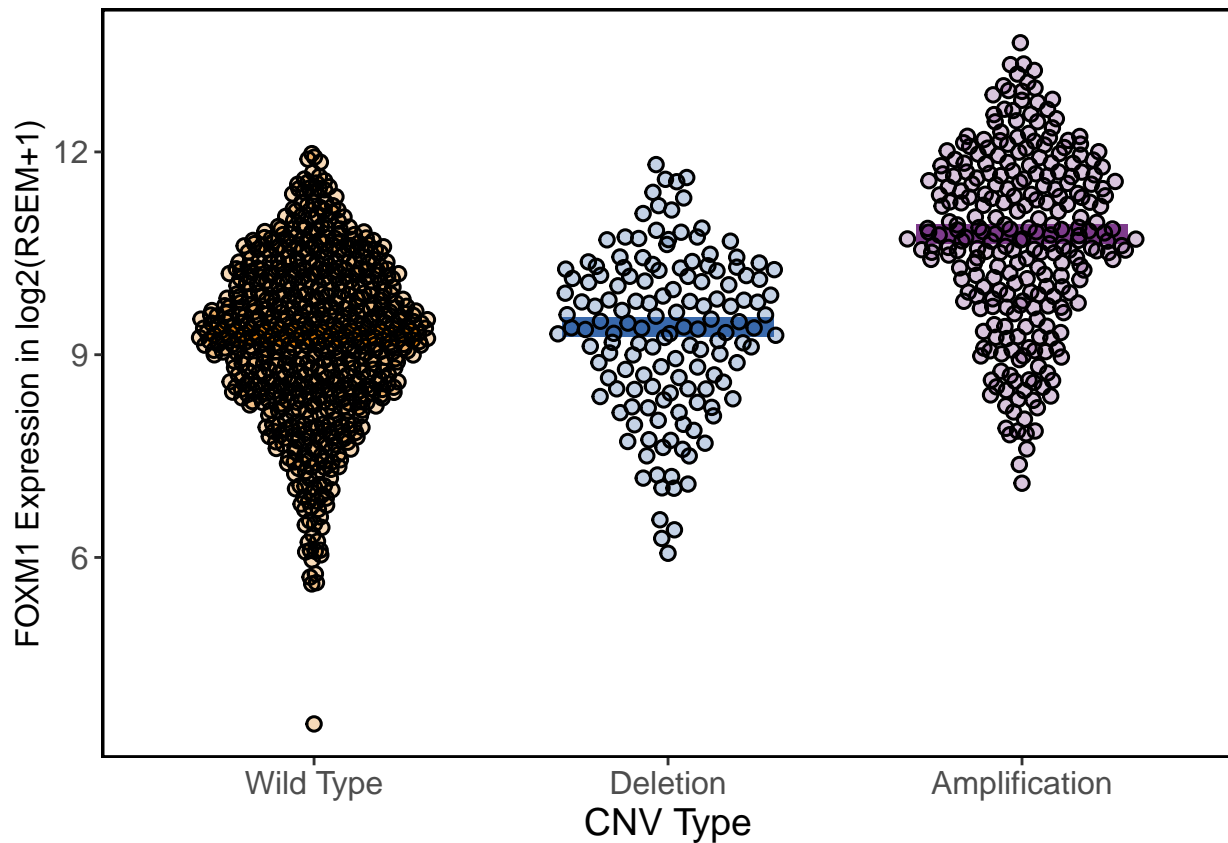


```
#MCM2 Expression
MCM2_Expr <- ggplot(Merged, aes(x =factor(MCM2.CNV, levels = c('Wild Type', 'Deletion', 'Amplification')
                                , y = MCM2, color = MCM2.CNV, fill = MCM2.CNV)) +
        scale_y_continuous() + labs(x="CNV Type", y='MCM2 Expression in log2(RSEM+1)')+
        scale_color_manual(values = my_pal, guide = "none") + scale_fill_manual(values = my_pal,
                                                                guide = "none")


MCM2_Expr<- MCM2_Expr + theme(axis.title.x =  element_text(size = 14), axis.title.y =
                               (element_text(size = 12)), axis.text = element_text(size = 12),   pane
                               element_rect(fill = 'NA', color = 'white'),panel.border = element_rec


MCM2_Expr <- MCM2_Expr +
```
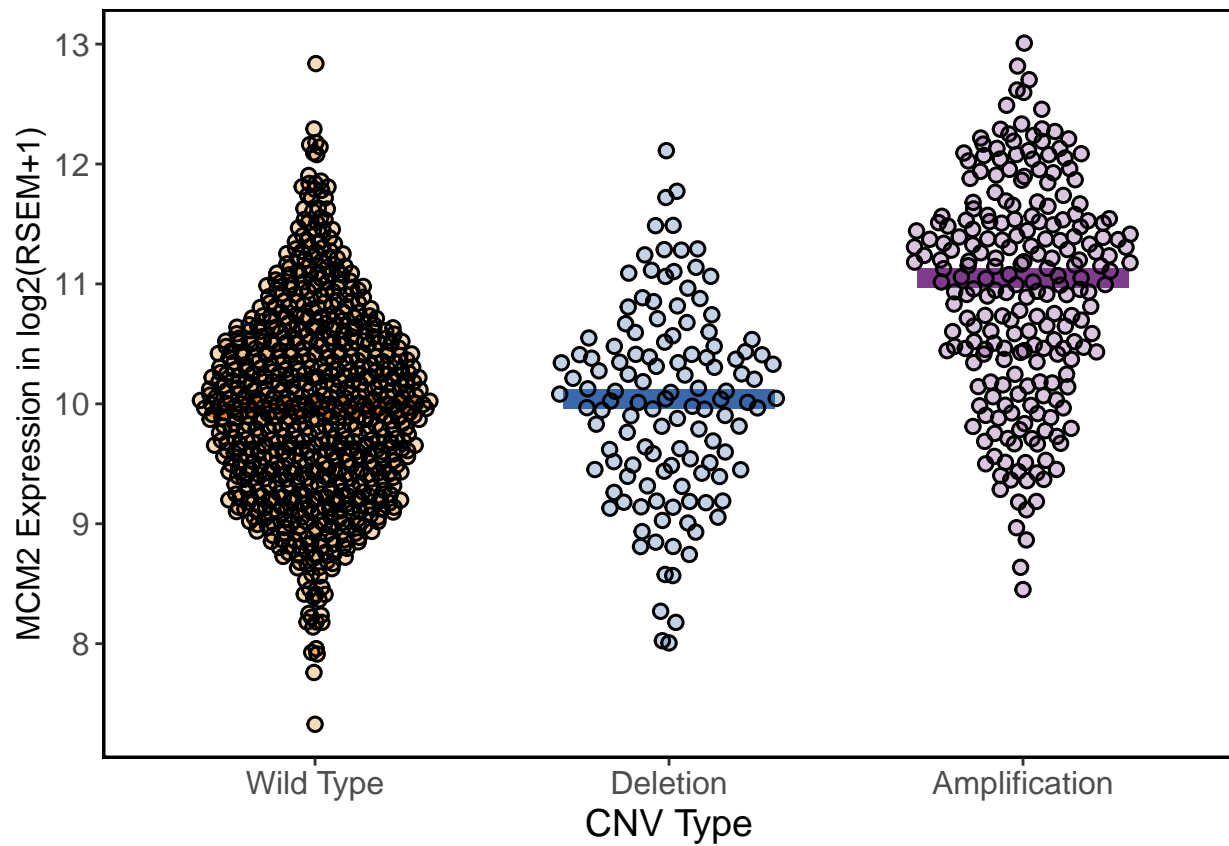
```
ggbeeswarm::geom_quasirandom(
  size = 2, width = .33, alpha = .3
) +
stat_summary(
  fun = median, geom = "point",
  shape = 95, size = 50
) +
ggbeeswarm::geom_quasirandom(
  size = 2, width = .33, shape = 1, color = "black", stroke = .8
)
```
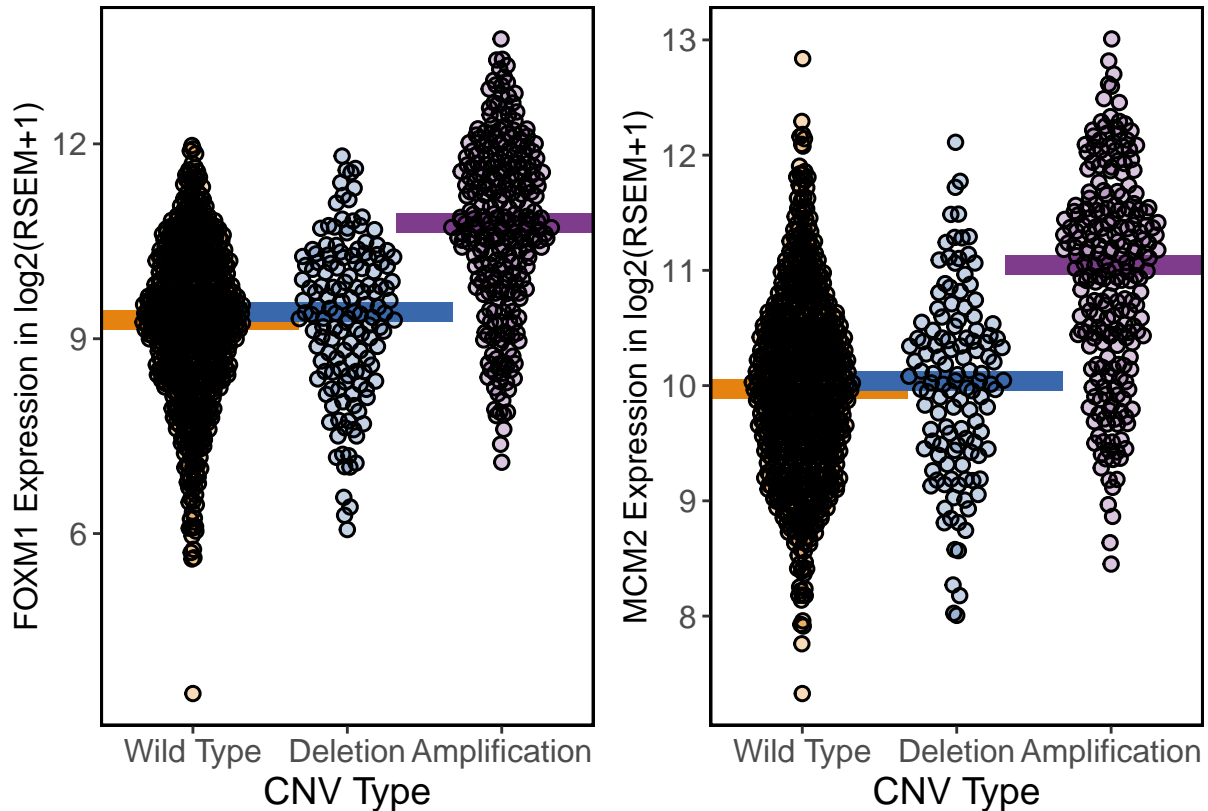
```
MCM2_Expr
```



```
FOXM1_Expr+MCM2_Expr
```

```
#Performing survival analysis
clinical_df<- read.table('data_clinical_patient.txt', header = T, sep = '\t')
clinical_df$PATIENT_ID<- gsub("-",".", as.character(clinical_df$PATIENT_ID))
Merged$PATIENT_ID<- gsub(".01","", as.character(Merged$PATIENT_ID))
cnv_surv<- left_join(Merged, clinical_df, by='PATIENT_ID')
cnv_surv_final<- cnv_surv[,c(1,6,7,36,37)]

#Omit blanks and NA
cnv_surv_final[cnv_surv_final=='']<- NA
cnv_surv_final<- na.omit(cnv_surv_final)

#create censored group
cnv_surv_final$censored<- ifelse(cnv_surv_final$OS_STATUS=='0:LIVING', FALSE, TRUE)


#Fitting model for survival analysis
fit1 <- survfit(Surv(OS_MONTHS, censored) ~ FOXM1_Sum, data = cnv_surv_final)
fit1
```
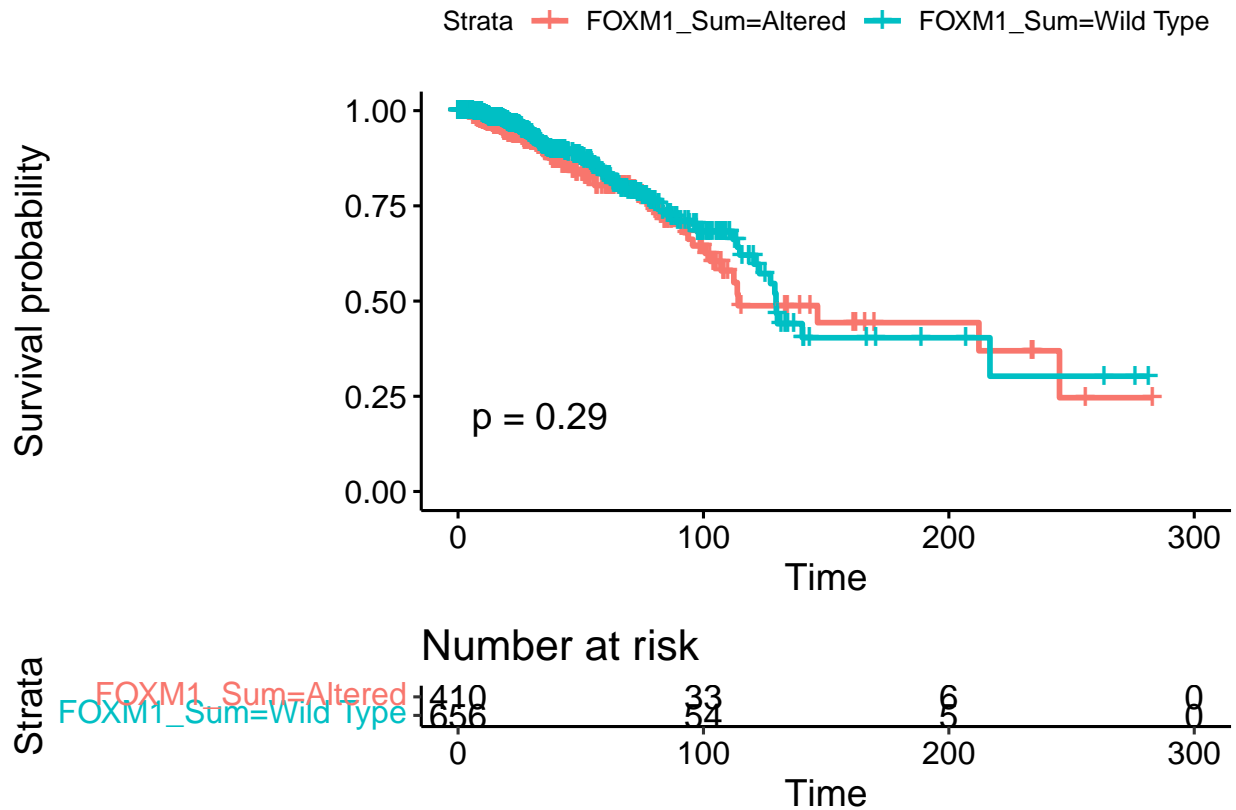
```
## Call: survfit(formula = Surv(OS_MONTHS, censored) ~ FOXM1_Sum, data = cnv_surv_final)
##
##                         n events median 0.95LCL 0.95UCL
## FOXM1_Sum=Altered     410     62    114     107      NA
## FOXM1_Sum=Wild Type   656     86    130     123      NA
```

```
FOXM1_Surv<-ggsurvplot(fit1,
          data = cnv_surv_final,
          pval = T,
          risk.table = T)
FOXM1_Surv
```


Strata + FOXM1_Sum=Altered + FOXM1_Sum=Wild Type

```
fit2 <- survfit(Surv(OS_MONTHS, censored) ~ MCM2_Sum, data = cnv_surv_final)
fit2
```

```
## Call: survfit(formula = Surv(OS_MONTHS, censored) ~ MCM2_Sum, data = cnv_surv_final)
##
##                       n events median 0.95LCL 0.95UCL
## MCM2_Sum=Altered     352     50    130     107      NA
## MCM2_Sum=Wild Type   714     98    130     115      NA
```

```
MCM2_Surv<-ggsurvplot(fit2,
                data = cnv_surv_final,
                pval = T,
                risk.table = T)
MCM2_Surv
```