TCGA Meth Data Analysis

Asad

5/8/2023

```
library(tidyverse)
```

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

## v ggplot2 3.4.2     v purrr   1.0.1
## v tibble  3.2.1     v dplyr   1.1.1
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.2     v forcats 0.5.1

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

```
library(dplyr)
library(ggplot2)
library(ggpubr)
library(wesanderson)
library(RColorBrewer)
library(survival)
library(survminer)
```

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

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

#Read data

```
setwd('D:/CancerData/TCGA BRCA cBioPortal/BRCA/BRCA Epigenomic')
#meth<-read.delim('data_methylation_hm27_hm450_merged.txt', header = T, sep = '\t')
#meth_gene<- filter(meth, NAME=='PSMC1'| NAME=='MCM2'| NAME=='BRCA1'| NAME=='BRCA2')
#write.csv(meth_gene, 'meth_gene.csv')
meth_gene<- read.csv('meth_gene.csv')
```

#Read clinical data

```
clin_data<-read.table('data_clinical_patient.txt', header = T, sep = '\t')
```

#Prepare datasets

```
meth_gene$PATIENT_ID<-gsub(".01",'', as.character(meth_gene$PATIENT_ID))
clin_data$PATIENT_ID<-gsub("-",'.', as.character(clin_data$PATIENT_ID))
```

#Merge

```
Merged<- left_join(meth_gene,clin_data, by='PATIENT_ID')
Merged_final<- Merged[,c(1:6,9:11,24,30,34,35)]
#write.csv(Merged_final, 'Merged_final.csv')
Merged_final<-read.csv('Merged_final.csv')
```

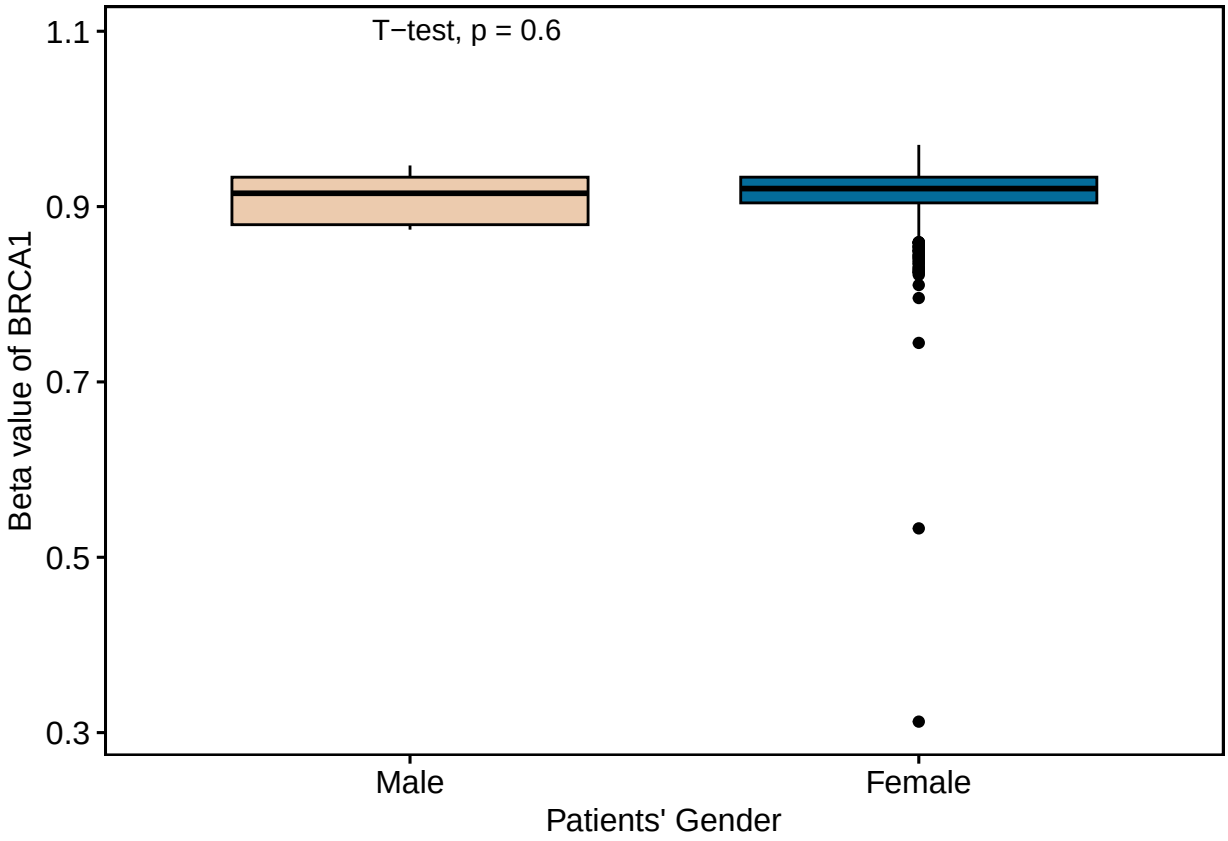
#Check methylation pattern in accordance with patients' gender

```
Merged_gender<- Merged_final %>% select('BRCA1', 'PSMC1', 'BRCA2', "MCM2", "SEX")
Merged_gender[Merged_gender=='']<-NA
Merged_gender<- na.omit(Merged_gender)
```

#Change color

```
col<- wes_palette(n=4, name='Darjeeling2')

BRCA1_gender<-ggboxplot(Merged_gender, x='SEX', y='BRCA1', fill = 'SEX', order =
                          c("Male", "Female"), palette = col)+xlab("Patients' Gender")+
                          ylab('Beta value of BRCA1')+stat_compare_means(method='t.test',
                          paired = F, label.y = 1.09)+border()
BRCA1_gender<- ggpar(BRCA1_gender, legend = 'none')
BRCA1_gender
```



##Check methylation pattern in accordance with patients' age

```
Merged_age<- Merged_final %>% select ('BRCA1', 'PSMC1', 'BRCA2', "MCM2", "AGE")
Merged_age[Merged_age==''']<-NA
Merged_age<- na.omit(Merged_age)
range(Merged_age$AGE)
```

```
## [1] 26 90
```

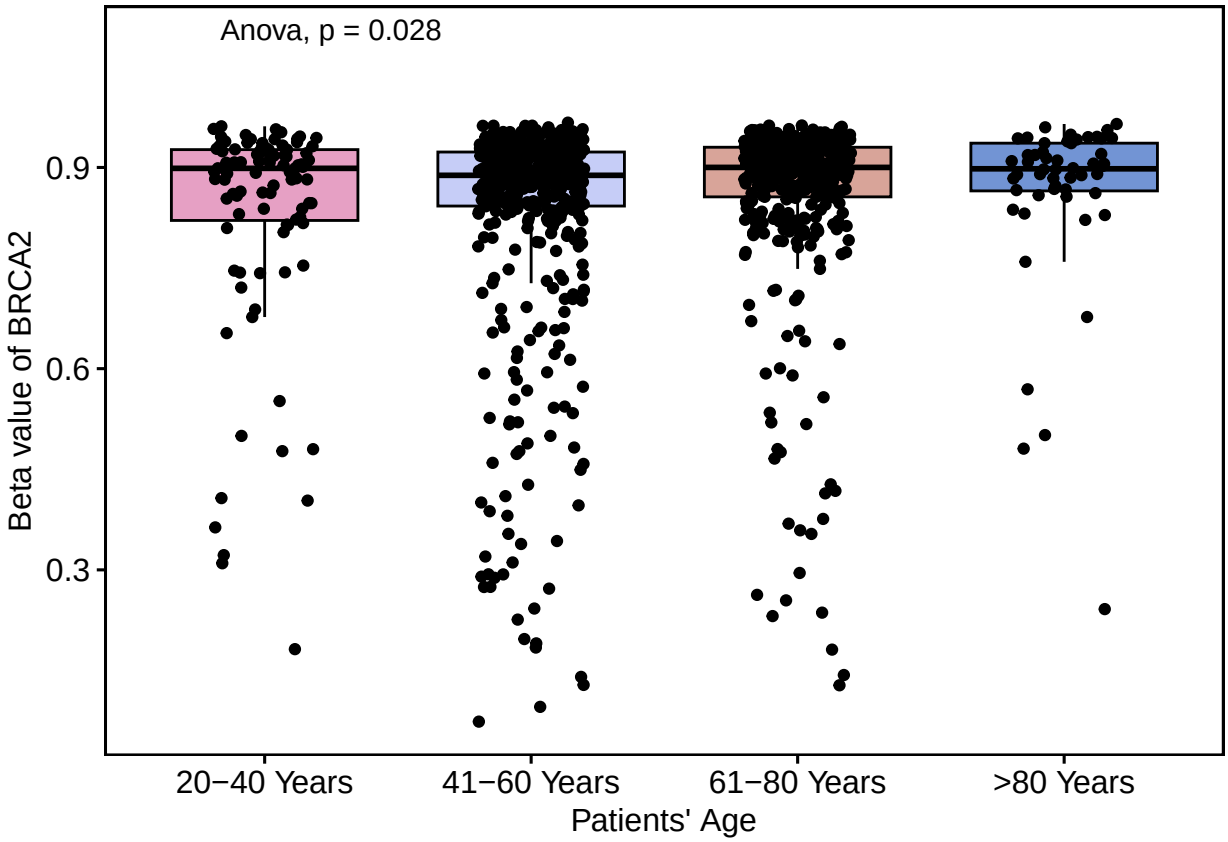
```
col2<- wes_palette(n=4, name='GrandBudapest2')
```

#Prepare age group

```
Merged_age["AGE_GROUP"] <- cut(Merged_age$AGE, c(19, 40, 60, 80, Inf),
                           c("20-40 Years", "41-60 Years", "61-80 Years", ">80 Years"), include.lowest=TR

BRCA2_age<-ggboxplot(Merged_age, x='AGE_GROUP', y='BRCA2', fill = 'AGE_GROUP', order =
                           c("20-40 Years", "41-60 Years", "61-80 Years", ">80 Years"),
                            palette = col2, add = 'jitter')+xlab("Patients' Age")+ ylab('Beta value of BR
                            stat_compare_means(method='anova',paired = F, label.y = 1.09)+border()

BRCA2_age<- ggpar(BRCA2_age, legend = 'none')
BRCA2_age
```

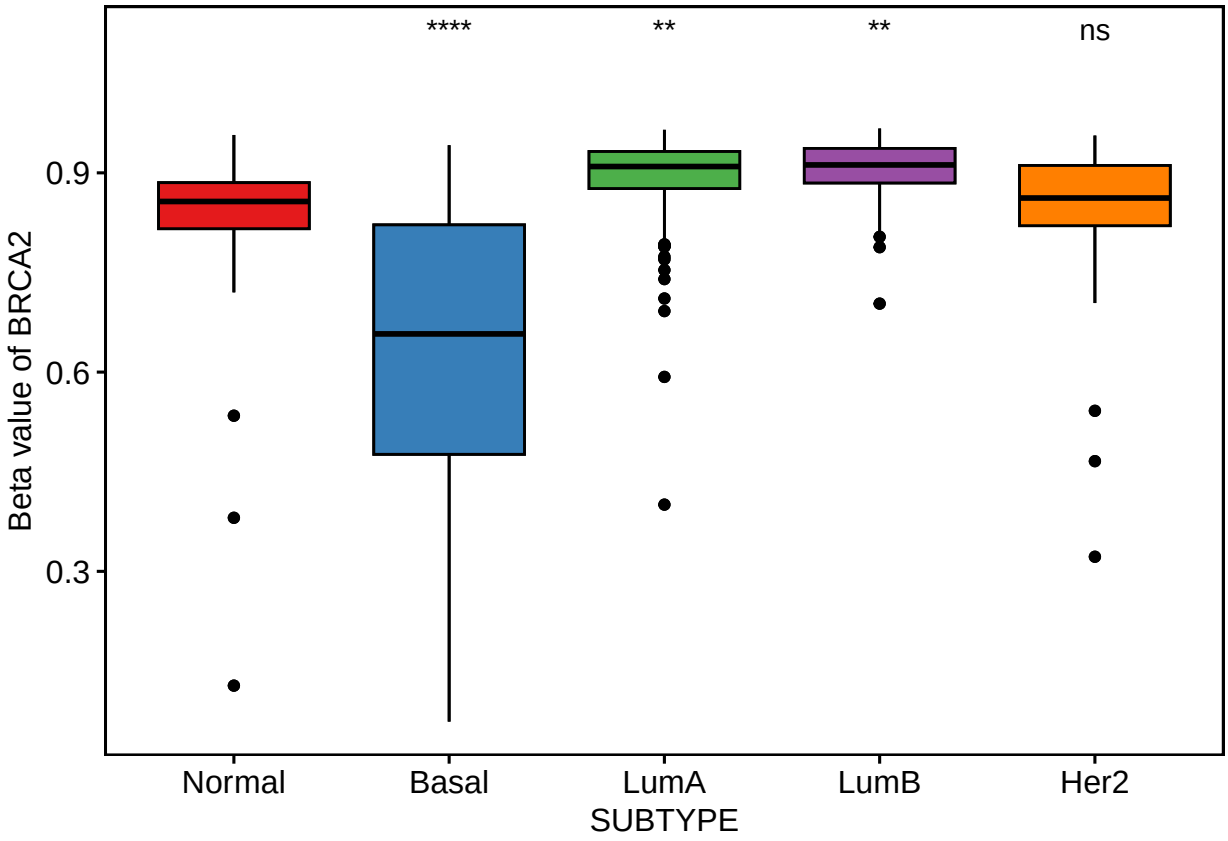


##Check methylation pattern in accordance with patients' age

```
Merged_sub<- Merged_final %>% select ('BRCA1', 'PSMC1', 'BRCA2', "MCM2", "SUBTYPE")
Merged_sub$SUBTYPE<- gsub("BRCA_","", as.character(Merged_sub$SUBTYPE))
Merged_sub[Merged_sub==''']<-NA
Merged_sub<- na.omit(Merged_sub)

BRCA2_Subtype<- ggboxplot(Merged_sub, x = "SUBTYPE", y = "BRCA2",
                          fill = "SUBTYPE",  palette = brewer.pal(n=9,name='Set1'),
                          order = c('Normal','Basal', 'LumA', 'LumB', 'Her2'), add = 'boxplot',
                          xlab = "SUBTYPE", ylab = 'Beta value of BRCA2') +
                         stat_compare_means(method = 't.test', label = 'p.signif',
                         ref.group = 'Normal', label.y = 1.1) + border()

BRCA2_Subtype<- ggpar(BRCA2_Subtype, legend = 'none')
BRCA2_Subtype
```



#Performing survival analysis

```
Surv<- Merged_final
x<- mean(unlist(Merged_final$BRCA2))
Surv$Group<- ifelse(Surv$BRCA2>=0.8441802, 'High Methylation', 'Low Methylation')
Surv$Censored<- ifelse(Surv$OS_STATUS=='0:LIVING',FALSE, TRUE)
Surv<- Surv[,c(5, 13:15)]
Surv[Surv==""]<- NA
Surv<- na.omit(Surv)
```

#Fitting model

```
fit <- survfit(Surv(OS_MONTHS, Censored) ~ Group, data = Surv)
fit
```

```
## Call: survfit(formula = Surv(OS_MONTHS, Censored) ~ Group, data = Surv)
##
##                          n events median 0.95LCL 0.95UCL
## Group=High Methylation 804    113    129     114     217
## Group=Low Methylation  260     34    245     245      NA
```

```
surv_pvalue(
  fit,
  data = Surv,
  method = "FH_p=1_q=1",
```

```
  test.for.trend = FALSE,
  combine = FALSE
)
```

```
##   variable   pval                          method pval.txt
## 1    Group 0.1766 Fleming-Harrington (p=1, q=1) p = 0.18
```

```
ggsurvplot(fit,
           data = Surv,
           risk.table = F,
           ggtheme = theme_bw(),
           palette = c("red", "blue"))
```