Supplementary

Appendix 1 Commands of R software

```
rm(list = ls())
setwd("D:\\bioinformatics\\Level 4\\L85_Codes\\work")
options("repos"= c(CRAN="https://mirrors.tuna.tsinghua.edu.cn/CRAN/"))
options(BioC_mirror="http://mirrors.ustc.edu.cn/bioc/")
if (!requireNamespace("BiocManager", quietly = TRUE))
install.packages("BiocManager")
BiocManager::install('tidyverse')
library(tidyverse)
options(stringsAsFactors = FALSE)
df_geo<-read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work\\GEO_R.csv',header = T,row.names = 1)
df_out<-read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work\\OUT_R1.csv',header = T,row.names = 1
df<-read.csv('F:\\bioinformatics\\Level 4\\L81_Codes\\Lesson1_Rawdata\\PartI_clinical_data.csv',header = T,row.names = 1)
df_out[1:5,1:5]
apply(df_geo, 2, class)
dim(df_out)
source('D:\\bioinformatics\\Level 4\\L81_Codes\\Lesson1_Function\\FindoutNA.R', encoding = "utf-8")
FindoutNA(df_geo)
FindoutNA(df_out)
df_geo_omit <- na.omit(df_geo)
dim(df_geo)
dim(df_geo_omit)
nrow(df_geo) - nrow(df_geo_omit)
rm(df_train)
rm(FindoutNA)
table(df_geo_omit$gender)
summary(df_geo_omit$age)
str(StepI_Rawdata)
str(df_geo_omit)
df_geo_omit$gender<-factor(df_geo_omit$gender,levels = c(0,1),labels = c('female','male'))
df_geo_omit$futime<-as.numeric(as.character(df_geo_omit$futime))
df_geo_omit$fustatus<-as.numeric(as.character(df_geo_omit$fustatus))
df_geo_omit$HB<-as.numeric(as.character(df_geo_omit$age))
str(df_out)
df_out$PLT<-as.numeric(as.character(df_out$PLT))
df_out$futime<-as.numeric(as.character(df_out$futime))
write.csv(df_geo_omit,'D:\\bioinformatics\\Level 4\\L85_Codes\\work\\GEO_Rdata.csv')
write.csv(df_out,'D:\\bioinformatics\\Level 4\\L85_Codes\\work\\OUT_Rdata.csv')
df_train <- read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work\\GEO_Rdata.csv',header = T,row.names = 1)
df_validation <- read.csv('D:\bioinformatics\Level 4\L85_Codes\work\\OUT_Rdata.csv',header = T,row.names = 1)

library(rms)
library(ResourceSelection)
library(dplyr)
df_omit_clear <- data.frame(Samples_ID = rownames(df_omit_clear),
df_omit_clear)
library(rms)
```


https://dx.doi.org/10.21037/tcr-23-1192

```
library(ResourceSelection)
library(dplyr)
library(caret)
library(ggpubr)
dir.create('F:\\bioinformatics\\Level 4\\L85_Codes\\work\\StepV_Compare_Results2')
source(file = 'F:\\bioinformatics\\Level 4\\L81_Codes\\Lesson1_Function\\GetTable_comparet.R', encoding = "utf-8")
setwd('F:\\bioinformatics\\Level 4\\L85_Codes\\work')
GetTable_compare(data1 = df_train,
data2 = df_validation,
df_name = 'Compare')
setwd('G:/Others_analysis/JLX_Nomogram/Lesson1/Part1')
rm(GetTable_compare)
df <- read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\Total3.csv',header = T,row.names = 1)
set.seed(43)
library(rms)
library(ResourceSelection)
library(dplyr)
df_omit_clear <- data.frame(Samples_ID = rownames(df), df)
library(rms)
library(ResourceSelection)
library(dplyr)
library(caret)
index_train <- createDataPartition(y = df$futime,
p = 0.5,
list = FALSE)
df_train <- df_omit_clear[index_train, ]
df_validation <- df_omit_clear[-index_train, ]
rm(index_train)
rm(df_omit_clear)
dir.create('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43')
source(file = 'D:\\bioinformatics\\Level 4\\L81_Codes\\Lesson1_Function\\GetTable_comparet.R', encoding = "utf-8")
setwd('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43')
GetTable_compare(data1 = df_train,
data2 = df_validation,
df_name = 'Compare')
```

LASSO regression analysis

```
library(glmnet)
library(survival)
library(tidyverse)
df_train_lasso <- read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_train_lasso.
    csv',header = T,row.names = 1)
for(i in names(dt)[c(1:9)]) {dt[,i] <- as.factor(dt[,i])}
x.factors <- model.matrix(~ dt$age+dt$diagnosis+dt$ASXL1+dt$CEBPA+dt$DNMT3A+dt$IDH2+dt$RUNX1+dt$TP53,dt)[,-1]
x <- as.matrix(data.frame(x.factors,dt[,10:11]))
y <- data.matrix(Surv(dt$time,dt$status))
dt$time<-as.numeric(dt$time)
```


https://dx.doi.org/10.21037/tcr-23-1192

```
dt<-df_train_lasso
str(dt)
fit <-glmnet(x.factors,y,family = "cox",alpha = 1)
plot(fit,label=T)
plot(fit,xvar="lambda",label=T)
fitcv <- cv.glmnet(x.factors,y,family="cox", alpha=1,nfolds=10)
plot(fitcv)
coef(fitcv, s="lambda.min")
expr_df<-dt%>%as.matrix()
coef.min=coef(fitcv,s="lambda.min")
active.min=which(coef.min!=0)
(lasso_geneids<-colnames(expr_df)[active.min])
df_train<-df_train[,-1]
df_validation<-df_validation[,-1]
df_train$fustatus<-as.numeric(as.character(df_train$fustatus))
str(df_train)
df_train$futime<-as.numeric(as.character(df_train$futime))
BaSurv<-Surv(time = df_train$futime,
event = df_train$fustatus)
Unicox<-function(x){
FML<-as.formula(paste0('BaSurv~',x))
GCox<-coxph(FML,data = df_train)
GSum<-summary(GCox)
HR<-round(GSum$coefficients[,2],2)
PValue<-round(GSum$coefficients[,5],3)
CI<-paste0(round(GSum$conf.int[,3:4],2),collapse = "-")
Unicox<-data.frame("characteristics"=x,
"Hazard Ratio"=HR,
"CI95"=CI,
"P Value"=PValue)
return(Unicox)
}
Unicox(colnames(df_train)[4])
VarNames<-colnames(df_train)
VarNames<-c("gender","age","diagnosis","WBC","HB","ASXL1","CEBPA","DNMT3A","EZH2","FLT3ITD","FLT3TKD","ID
    H1","IDH2","NPM1","RUNX1","TET2","TP53")
UniVar<-lapply(VarNames,Unicox)
UniVar<-ldply(UniVar,data.frame)
library(plyr)
GetFactors_uni<-UniVar$characteristics[which(UniVar$P.Value<0.2)] %>% as.character()
```

Multifactor COX regression analysis

```
fml<-as.formula(paste0('BaSurv~',paste0(GetFactors_uni,collapse = '+')))
MultiCox<-coxph(fml,data = df_train)
MultiSum<-summary(MultiCox)
MHR<-round(MultiSum$coefficients[,2],2)
setwd('G:/Others_analysis/JLX_Nomogram/Lesson1/Part1')
```


https://dx.doi.org/10.21037/tcr-23-1192

```
library(ggpubr)
rm(GetTable_compare)
write.csv(df_train,'D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_train.csv')
write.csv(df_validation,'D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_validation.csv')
```

nomogram

```
Final<-read.csv(file = "D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\Final.csv",header =
    T,row.names = 1,encoding = "UTF-8")
(Final_GetFactors <- Final$characteristics[which(Final$P.Value.y < 0.05)] %>% as.character())
# save(GetFactors,file ='Chr3_Univariate_Cox.RData')
fml<-as.formula(paste0('BaSurv~',paste0(Final_GetFactors,collapse = "+")))
MultiCox<-coxph(fml,data = df_train)
MultiSum<-summary(MultiCox)
Final_GetFactors<-c("age","diagnosis","DNMT3A","IDH2","TP53")
dd<-datadist(df_train)
options(datadist="dd")
BaSurv<-Surv(time = df_train$futime,event = df_train$fustatus)
fml<-as.formula(paste0('BaSurv~',paste0(Final_GetFactors,collapse = "+")))
f<-cph(fml,x=T,y=T,surv=T,data = df_train)
surv<-Survival(f)
nom<-nomogram(f,
fun = list(function(x)surv(365,x),
function(x)surv(365*2,x),
function(x)surv(365*3,x)),
lp=T,funlabel=c("1-year survival","2-year survival","3-year survival"),
maxscale=100,
fun.at=seq(0.1,0.9,0.1))
plot('normgram.plot',width = 12,height=12,onefile = FALSE)
plot(nom,cex.var=2,cex.axis=1.5,lwd=10,xfrac=0.5,tcl=0.5)
dev.off()
```

dynamic nomogram

```
install.packages('shinyPredict')
library(shinyPredict)
Cox_nomo<-coxph(Surv(futime, fustatus)~ age+diagnosis+DNMT3A+IDH2+TP53,
data = df_train,model = F,y=F)
shinyPredict(models =list("model 1"=Cox_nomo),
path ="D://bioinformatics//Level 4//L85_Codes//work//StepV_Compare_Results4seed49//gyc2415940441",
data = df_train[,c(2:4,13:14,18:20)],
title ="Dynamic nomogram",
shinytheme ="paper")
str(df_train)
install.packages('rsconnect')
library(rsconnect)
rsconnect::setAccountInfo(name='gyc2415940441',
token='F05C38E1C858DFF6C8BD718560615845',
```


https://dx.doi.org/10.21037/tcr-23-1192

```
secret='1mB2GC0DJQuHq5rx2TvMWnN2oJBqCsPOVdIJLC8e')
if(!require(rsconnect)) install.packages("rsconnect")
rsconnect::setAccountInfo(name='gyc2415940441',
token='72B83328FD03A5E06AABF2554AD27C73',
secret='DeKq8KzhSmIibeWknftpJB7A1f6JL2iebZShWqHs')
```

ROC curve

```
df_train_ROC <- read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_train_ROC.
    csv',header = T)
install.packages('timeROC')
library(rms)
library(timeROC)
data<-df_train_ROC
ROC <- timeROC(T = data$futime/(365.5/12),
delta = as.numeric(data$fustatus),
marker = data$riskscore,
cause = 1,
weighting = "marginal",
times = c(12, 24, 36),
iid = T)
ROC
confint(ROC)$CI_AUC
{
plot(ROC, time=12*1, lwd=2,col = "blue", add = F, title = F)
plot(ROC, time=12*2, lwd=2,col = "red", add = T)
plot(ROC, time=12*3, lwd=2,col = "black", add = T)
legend(x=0.5,y=0.25, text.width=1,
x.intersp=0.6,y.intersp=0.5,lty = 1, cex = 1,bty='n',
col = c("blue", "red", "black"),
legend = c("1y AUC:0.755",
"2y AUC:0.745",
"3y AUC:0.757"))
}

dev.new()
```

Discrimination analysis

```
library(rms)
df_train <- read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_train.csv',header =
    T,row.names = 1)
df_train$futime<-as.numeric(as.character(df_train$futime))
df_train$fustatus<-as.numeric(as.character(df_train$fustatus))
str(df_train)
Final_GetFactors<-c("age","diagnosis","DNMT3A","IDH2","TP53")
BaSurv<-Surv(time = df_train$futime,event = df_train$fustatus)
fml<-as.formula(paste0('BaSurv~',paste0(Final_GetFactors,collapse = "+")))
```


https://dx.doi.org/10.21037/tcr-23-1192

```
dd<-datadist(df_train)
options(datadist="dd")
f<-cph(fml,x=T,y=T,surv=T,data = df_train)
validate(f,method = "boot",B=1000,dxy = T)
c_index<-rcorrcens(Surv(futime,fustatus)~predict(f),data = df_train)
index<-1-c_index[1]%>%round(.,3)
low95CI<-c(index-c_index[4]/2)%>%round(.,3)
up95CI<-c(index+c_index[4]/2)%>%round(.,3)
sink('c_index.txt')
(cindex_df<-data.frame(c_index=index,low95CI=low95CI,up95CI=up95CI))
sink()
```

calibration analysis

```
f1<-cph(fml,x=T,y=T,surv = T,data = df_train,time.inc = 365)
install.packages("rms")
dev.off()
library(rms)
cal1<-calibrate(f1,
cmethod = "KM",
method = "boot",
u=365,
m=nrow(df_train)/3,
B=1000)
pdf('cal11.pdf',width = 12,height = 12,onefile = F)
plot(cal1,
lwd=2,
lty=0,
conf.int=F,subtitles = FALSE,
riskdist = FALSE,par.corrected=list(col='white'),
errbar.col=c("#1159AC"),
xlab = "Nomogram-Predicted Probability",
ylab = "Actual survival",
xlim=c(0,1),ylim=c(0,1),
cex.lab=1.0,cex.axis=1,cex.main=1.2,cex.sub=0.6,add = F)
lines(cal1[,c('mean.predicted',"KM")],
type = 'b',
lwd=3,
pch=16,
col=c("#548C00"))
mtext("")
box(lwd=2)
abline(0,1,lty=3,
lwd=2,
col=c("gray66")
)
dev.new()
```


https://dx.doi.org/10.21037/tcr-23-1192

```
f2<-cph(fml,x=T,y=T,surv = T,data = df_train,time.inc = 365*2)
cal2<-calibrate(f2,
cmethod = "KM",
method = "boot",
u=365*2,
m=nrow(df_train)/3,
B=1000)
plot(cal2,
lwd=2,
lty=0,
conf.int=F,subtitles = FALSE,
riskdist = FALSE,par.corrected=list(col='white'),
errbar.col=c("#1159AC"),
xlab = "Nomogram-Predicted Probability of 1-year OS",
ylab = "Actual 1-year OS(Proportion)",
cex.lab=1.0,cex.axis=1,cex.main=1.2,cex.sub=0.6,add = T)
lines(cal2[,c('mean.predicted',"KM")],
type = 'b',
lwd=3,
pch=16,
col=c("#A23400"))
mtext("")
box(lwd=2)
abline(0,1,lty=3,
lwd=4,
col=c("#224444")
)

f3<-cph(fml,x=T,y=T,surv = T,data = df_train,time.inc = 365*3)
cal3<-calibrate(f3,
cmethod = "KM",
method = "boot",
u=365*3,
m=nrow(df_train)/3,
B=1000)
plot(cal3,
lwd=2,
lty=0,
conf.int=F,subtitles = FALSE,
riskdist = FALSE,par.corrected=list(col='white'),
errbar.col=c("#1159AC"),
xlab = "Nomogram-Predicted Probability of 1-year OS",
ylab = "Actual 1-year OS(Proportion)",
cex.lab=1.0,cex.axis=1,cex.main=1.2,cex.sub=0.6,add = T)
lines(cal3[,c('mean.predicted',"KM")],
type = 'b',
lwd=3,
pch=16,
```


https://dx.doi.org/10.21037/tcr-23-1192

```
col=c("#2166AC"))
mtext("")
box(lwd=2)
abline(0,1,lty=3,
lwd=2,
col=c("#224444")
)
legend("bottomright", bty = 'n',
legend=c("1-year", "2-year","3-year"),
col=c("#548C00", "#A23400","#2166AC"),
lwd=2,plot = T)
```

DCA curves

```
install.packages("htmltools")
library(ggDCA)
library(foreign)
dca_train<-dca(f)
f<-cph(Surv(futime,fustatus)~age+diagnosis+DNMT3A+IDH2+TP53,df_train)
install.packages("ggprism")
library(ggprism)
ggplot(dca_train,linetype =F,lwd = 1.2)+
theme_classic()+
theme_prism(base_size =17)+
theme(legend.position="top")+
theme(axis.line.x=element_line(size=0.5))+
theme(axis.line.y=element_line(size=0.5))+
scale_x_continuous(
limits = c(0, 1),
guide = "prism_minor") +
theme(axis.ticks.x=element_line(size=0.5))+
theme(axis.ticks.y=element_line(size=0.5))+
theme(axis.text.x = element_text(face="plain",size=10))+
theme(axis.text.y = element_text(face="plain",size=10))+
scale_y_continuous(
limits = c(-0.03, 0.4),
guide = "prism_minor")+
scale_colour_prism(
palette = "candy_bright",
name = "Cylinders",
label = c("nomogram", "ALL", "None"))+
theme(legend.position = c(0.8,0.8))+
theme(text = element_text(size = 16))+
theme(axis.title = element_text(face="plain",size = 12))
```

Risk scores

```
Final_GetFactors<-c("ELN","ELN")
```


https://dx.doi.org/10.21037/tcr-23-1192

```
fml<-as.formula(paste0('BaSurv~',paste0(Final_GetFactors,collapse = '+')))
df_geo<-df_geo_ELN
MultiCox<-coxph(fml,data = df_geo)
MultiSum<-summary(MultiCox)
index.min<-MultiSum$coefficients[,1]
(index.min<-as.numeric(index.min))
signature<-as.matrix(subset(df_train,select = Final_GetFactors))%*%as.matrix(exp(index.min))
library(rms)
install.packages('ggrisk')
library(ggrisk)
library(pheatmap)
library(ggplot2)
library(ggplotify)
library(cowplot)
summary(MultiCox,data=df_geo)
riskscore<-predict(MultiCox,type = "risk",df_geo)
names(riskscore)=rownames(df_geo)
write.csv(riskscore,"df_geo_ROC")
fp<-riskscore
phe<-outer
fp_dftrain=data.frame(patientid=1:length(fp),fp=as.numeric(sort(fp)))
fp_dftrain$riskgroup=ifelse(fp_dftrain$fp>=1.637178,'high','low')
sur_dat=data.frame(patientid=1:length(fp),
time=phe[names(sort(fp)),'futime'],
event=phe[names(sort(fp)),'fustatus'])
sur_dat$event=ifelse(sur_dat$event==0,'alive','dead')
sur_dat$event=factor(sur_dat$event,levels = c("dead","alive"))
exp_dat=dat[names(sort(fp)),(ncol(dat)-7):ncol(dat)]
library(ggplot2)
p1=ggplot(fp_dftrain,aes(x=patientid,y=fp))+geom_point(aes(color=riskgroup))+
scale_colour_manual(values = c("red3","blue3"))+
theme_bw()+labs(x="",y="Risk score")+
geom_vline(xintercept=sum(fp_dftrain$riskgroup=="low"),colour="black", linetype="dotted",linewidth=0.8)
p1
p2=ggplot(sur_dat,aes(x=patientid,y=time))+geom_point(aes(col=event))+theme_bw()+
scale_colour_manual(values = c("red3","blue3"))+
labs(x="Patient ID",y="Survival time(year)")+
geom_vline(xintercept=sum(fp_dftrain$riskgroup=="low"),colour="black", linetype="dotted",size=0.8)
p2
p1/p2
library(ggplot2)
library(survminer)
library(survival)
df_train_ROC <- read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_train_ROC.
    csv',header = T)
df_train_riskscore<-cbind(df_train,df_train_ROC[,3])
colnames(df_train_riskscore)[20]='riskscore'
write.csv(df_train_riskscore,'D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_train_
```

 https://dx.doi.org/10.21037/tcr-23-1192

```
    riskscore.csv')
df<-df_train_riskscore
df$riskscore_by2<-ifelse(df$riskscore>median(df$riskscore),'High-score','Low-score')
fit<-survfit(Surv(futime/30,fustatus)~riskscore_by2,data = df)
p<-ggsurvplot(fit,conf.int = T,pval = T,risk.table = T,
legend.labs=c('High-score','Low-score'),
legend.title='',palette = c("red3","blue3"),
risk.table.height=0.3,
break.time.by=12,xlab="Time(months)")
```

elderly vs young

train young

```
df_geo_riskscore <- read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_geo_
    riskscore.csv',header = T,row.names = 1)
df_geo<-rbind(df_train,df_validation)
df_geo_ROC<-rbind(df_train_ROC,df_validation_ROC)
df_geo <- read.csv('D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\Total3.csv',header = T,row.names = 1)
df_geo_riskscore<-cbind(df_geo,df_geo_ROC[,3])
colnames(df_geo_riskscore)[20]='riskscore'
write.csv(df_geo_riskscore,'D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\StepV_Compare_Results4seed43\\df_geo_
    riskscore.csv')
df<-df_geo_riskscore
df_young$riskscore_by2<-ifelse(df_young$riskscore>median(df_young$riskscore),'High-score','Low-score')
fit<-survfit(Surv(futime/30,fustatus)~riskscore_by2,data = df_young)
p_young<-ggsurvplot(fit,conf.int = T,pval = T,risk.table = T,
legend.labs=c('High-score','Low-score'),
legend.title='',palette = c("red3","blue3"),
risk.table.height=0.3,title="Young AML",
break.time.by=12,xlab="Time(months)")
p_young

df_young<-df[df$age=="adult",]
```

model comparison

```
library(ggDCA)
library(foreign)
dca_train<-dca(f)
df_outer<-read.csv("D:\\bioinformatics\\Level 4\\L85_Codes\\work1\\outer.csv",header = T, row.names = 1)
dd<-datadist(df_outer)
options(datadist="dd")
f<-cph(Surv(futime,fustatus)~age+diagnosis+DNMT3A+IDH2+TP53,df_outer)
f1<-cph(Surv(futime,fustatus)~ELN+ELN,df_outer)
install.packages("ggprism")
library(ggprism)
dt=dca(f,f1)
```


https://dx.doi.org/10.21037/tcr-23-1192

```
ggplot(dt,linetype =F,lwd = 1.2)+
theme_classic()+
theme_prism(base_size =17)+
theme(legend.position="top")+
theme(axis.line.x=element_line(size=0.5))+
theme(axis.line.y=element_line(size=0.5))+
scale_x_continuous(
limits = c(0, 1),
guide = "prism_minor") +
theme(axis.ticks.x=element_line(size=0.5))+
theme(axis.ticks.y=element_line(size=0.5))+
theme(axis.text.x = element_text(face="plain",size=10))+
theme(axis.text.y = element_text(face="plain",size=10))+
scale_y_continuous(
limits = c(-0.03, 0.4),
guide = "prism_minor")+
scale_colour_prism(
palette = "candy_bright",
name = "Cylinders",
label = c("nomogram","ELN","ALL","None"))+
theme(legend.position = c(0.8,0.8))+
theme(text = element_text(size = 16))+
theme(axis.title = element_text(face="plain",size = 12))
```


https://dx.doi.org/10.21037/tcr-23-1192