R code for the “cor.test”

library(limma)

corFilter=0.4

pvalueFilter=0.001

setwd("")

rt=read.table("lncRNA.txt", header=T, sep="\t", check.names=F)

rt=as.matrix(rt)

rownames(rt)=rt[,1]

exp=rt[,2:ncol(rt)]

dimnames=list(rownames(exp),colnames(exp))

data=matrix(as.numeric(as.matrix(exp)),nrow=nrow(exp),dimnames=dimnames)

data=avereps(data)

data=data[rowMeans(data)>0.1,]

group=sapply(strsplit(colnames(data),"\\-"),"[",4)

group=sapply(strsplit(group,""), "[", 1)

group=gsub("2","1",group)

lncRNA=data[,group==0]

conNum=length(group[group==1])

treatNum=length(group[group==0])

sampleType=c(rep(1,conNum), rep(2,treatNum))

rt1=read.table("EMTGeneExp.txt", header=T, sep="\t", check.names=F)

rt1=as.matrix(rt1)

rownames(rt1)=rt1[,1]

exp1=rt1[,2:ncol(rt1)]

dimnames1=list(rownames(exp1),colnames(exp1))

EMT=matrix(as.numeric(as.matrix(exp1)), nrow=nrow(exp1), dimnames=dimnames1)

EMT=avereps(EMT)

EMT=EMT[rowMeans(EMT)>0.1,]

group=sapply(strsplit(colnames(EMT),"\\-"),"[",4)

group=sapply(strsplit(group,""),"[",1)

group=gsub("2","1",group)

EMT=EMT[,group==0]

outTab=data.frame()

for(i in row.names(lncRNA)){

if(sd(lncRNA[i,])>0.1){

test=wilcox.test(data[i,] ~ sampleType)

if(test$p.value<0.05){

for(j in row.names(EMT)){

x=as.numeric(lncRNA[i,])

y=as.numeric(EMT[j,])

corT=cor.test(x,y)

cor=corT$estimate

pvalue=corT$p.value

if((cor>corFilter) & (pvalue<pvalueFilter)){

outTab=rbind(outTab,cbind(EMT=j,lncRNA=i,cor,pvalue,Regulation="postive"))

}

if((cor< -corFilter) & (pvalue<pvalueFilter)){

outTab=rbind(outTab,cbind(EMT=j,lncRNA=i,cor,pvalue,Regulation="negative"))

}

}

}

}

}

write.table(file="corResult.txt",outTab,sep="\t",quote=F,row.names=F)

EMTLncRNA=unique(as.vector(outTab[,"lncRNA"]))

EMTLncRNAexp=data[EMTLncRNA,]

EMTLncRNAexp=rbind(ID=colnames(EMTLncRNAexp), EMTLncRNAexp)

write.table(EMTLncRNAexp,file="EMTLncExp.txt",sep="\t",quote=F,col.names=F)

R code for figure Supplement Figure S1 and Figure 2A-B

library(survival)

pFilter=0.05

setwd("D:\\ ")

rt=read.table("expTime.txt", header=T, sep="\t", check.names=F, row.names=1)

rt$futime=rt$futime/365

outTab=data.frame()

sigGenes=c("futime","fustat")

for(gene in colnames(rt[,3:ncol(rt)])){

cox=coxph(Surv(futime, fustat) ~ rt[,gene], data = rt)

coxSummary = summary(cox)

coxP=coxSummary$coefficients[,"Pr(>|z|)"]

if(coxP<pFilter){

sigGenes=c(sigGenes,gene)

outTab=rbind(outTab,

cbind(gene=gene,

HR=coxSummary$conf.int[,"exp(coef)"],

HR.95L=coxSummary$conf.int[,"lower .95"],

HR.95H=coxSummary$conf.int[,"upper .95"],

pvalue=coxP) )

}

}

write.table(outTab,file="uniCox.txt",sep="\t",row.names=F,quote=F)

surSigExp=rt[,sigGenes]

surSigExp=cbind(id=row.names(surSigExp),surSigExp)

write.table(surSigExp,file="uniSigExp.txt",sep="\t",row.names=F,quote=F)

bioForest=function(coxFile=null, forestFile=null){

rt <- read.table(coxFile, header=T, sep="\t", check.names=F, row.names=1)

gene <- rownames(rt)

hr <- sprintf("%.3f",rt$"HR")

hrLow <- sprintf("%.3f",rt$"HR.95L")

hrHigh <- sprintf("%.3f",rt$"HR.95H")

Hazard.ratio <- paste0(hr,"(",hrLow,"-",hrHigh,")")

pVal <- ifelse(rt$pvalue<0.001, "<0.001", sprintf("%.3f", rt$pvalue))

pdf(file=forestFile, width=6.6, height=4.5)

n <- nrow(rt)

nRow <- n+1

ylim <- c(1,nRow)

layout(matrix(c(1,2),nc=2),width=c(3,2.5))

xlim = c(0,3)

par(mar=c(4,2.5,2,1))

plot(1,xlim=xlim,ylim=ylim,type="n",axes=F,xlab="",ylab="")

text.cex=0.8

text(0,n:1,gene,adj=0,cex=text.cex)

text(1.5-0.5\*0.2,n:1,pVal,adj=1,cex=text.cex);text(1.5-0.5\*0.2,n+1,'pvalue',cex=text.cex,font=2,adj=1)

text(3.1,n:1,Hazard.ratio,adj=1,cex=text.cex);text(3.1,n+1,'Hazard ratio',cex=text.cex,font=2,adj=1)

par(mar=c(4,1,2,1),mgp=c(2,0.5,0))

xlim = c(0,max(as.numeric(hrLow),as.numeric(hrHigh)))

plot(1,xlim=xlim,ylim=ylim,type="n",axes=F,ylab="",xaxs="i",xlab="Hazard ratio")

arrows(as.numeric(hrLow),n:1,as.numeric(hrHigh),n:1,angle=90,code=3,length=0.05,col="darkblue",lwd=2.5)

abline(v=1,col="black",lty=2,lwd=2)

boxcolor = ifelse(as.numeric(hr) > 1, "red", "green")

points(as.numeric(hr), n:1, pch = 15, col = boxcolor, cex=1.5)

axis(1)

x=as.matrix(train[,c(3:ncol(train))])

y=data.matrix(Surv(train$futime,train$fustat))

fit <- glmnet(x, y, family = "cox", maxit = 1000)

cvfit <- cv.glmnet(x, y, family="cox", maxit = 1000)

coef <- coef(fit, s = cvfit$lambda.min)

index <- which(coef != 0)

actCoef <- coef[index]

lassoGene=row.names(coef)[index]

geneCoef=cbind(Gene=lassoGene, Coef=actCoef)

if(nrow(geneCoef)<2){next}

pdf("lambda.pdf")

plot(fit, xvar = "lambda", label = TRUE)

dev.off()

pdf("cvfit.pdf")

plot(cvfit)

abline(v=log(c(cvfit$lambda.min,cvfit$lambda.1se)),lty="dashed")

dev.off()

}

R code for figure 2C-D

library(reshape2)

library(ggplot2)

library(scales)

library(cowplot)

setwd("")

data <- read.table("risk.train.txt", row.names = 1) #去除第一行第一列的id设置为空白

rs <- data$riskscore

names(rs) <- rownames(data)

bestvars <- read.table("signaturegenes.txt")$V1 #signature基因名

rs <- data$riskScore

names(rs) <- rownames(data)

rs\_data <- data.frame(x=1:length(rs),rs=as.numeric(sort(rs)))

rs\_data$Risk <- ifelse(rs\_data$rs>=median(rs\_data$rs), "High-risk", "Low-risk")

head(rs\_data)

surv\_data <- data.frame(x=1:length(rs),

t=data[names(sort(rs)),'futime'],

s=data[names(sort(rs)),'fustat'])

surv\_data$Status <- as.factor(ifelse(surv\_data$s==0,'Alive','Death'))

head(surv\_data)

exp\_data <- data[names(sort(rs)),which(colnames(data) %in% bestvars)]

plot.A <- ggplot(rs\_data, aes(x=x,y=rs))+

geom\_point(aes(col=Risk),size=0.5)+

scale\_color\_manual(labels=c("High-risk","Low-risk"),

#guide\_legend(guide = NULL),

name="Risk score", values =c("#DC0000FF", "#00A087FF")) +

geom\_segment(aes(x = sum(rs\_data$Risk=="Low-risk"),

y = 0,

xend = sum(rs\_data$Risk=="Low-risk"),

yend = max(rs\_data$rs)), linetype="dashed", size = 0.6)+

#geom\_segment(aes(x=0,y=median(rs\_data$rs),

# xend=nrow(rs\_data),

# yend=median(rs\_data$rs)),linetype="dashed", size = 0.3)+

#geom\_text(aes(x=sum(rs\_data$Risk=="Low-risk")/2,

# y=median(rs\_data$rs)+8,

# label=paste0("Cutoff: ",round(median(rs\_data$rs),3))),

# col ="black",size = 4,alpha=0.8)+

theme(axis.title.x=element\_blank()) +

scale\_x\_continuous(limits = c(0,NA),expand = c(0,0)) +

labs(y="Risk score",x="",fill="Risk") +

#scale\_colour\_discrete(name="Risk scores") +

theme\_classic() +

theme(axis.ticks.x=element\_blank(),

axis.line = element\_blank(),

axis.text.x=element\_blank())

plot.B <- ggplot(surv\_data,aes(x=x,y=t))+

geom\_point(aes(col=Status),size=0.5)+

geom\_vline(aes(xintercept=sum(rs\_data$Risk=="Low-risk")),size=0.6,linetype="dashed")+

scale\_x\_continuous(limits = c(0,NA),expand = c(0,0))+

scale\_color\_manual(labels=c("Alive","Dead"),

values =c("#00A087FF","#DC0000FF"))+

labs(y="OS(years)",x="")+

theme\_classic()+

theme(axis.ticks.x=element\_blank(),

axis.line = element\_blank(),

axis.text.x=element\_blank())

tmp <- t(scale(exp\_data))

tmp[tmp > 1] = 1

tmp[tmp < -1] = -1

reorder\_cormat <- function(cormat){

dd <- dist(cormat)

hc <- hclust(dd,method = "average")

cormat <-cormat[hc$order,]

}

tmp1 <- reorder\_cormat(tmp)

tmp1 <- rbind(tmp1,ifelse(rs\_data$Risk=="Low-risk",-1.5,1.5))

tmp.m <- melt(tmp1)

p2 <-ggplot(tmp.m, aes(Var2, Var1),size=0.5) +

geom\_tile(aes(fill = value))

plot.C <- p2 + scale\_fill\_gradient2(name="Genes\nexpression", low="#00A087FF", high="#DC0000FF", mid="white") +

labs(x = "", y = "")+

theme\_classic()+

theme(legend.title = element\_text(size = 12), legend.position = "right",

axis.line = element\_blank(),

axis.ticks=element\_blank(),

axis.text.x=element\_blank())

plot\_grid(plot.A, plot.B, plot.C,

labels = c("B", "",""), # 或者按顺序标注ABC

rel\_heights = c(1,1,1), # 3个图的比例

#label\_x=0,

#label\_y=1,

align = 'v',ncol = 1, axis="lr", scale = c(1,1,1), greedy = F)

ggsave("cox.pdf", width = 7, height = 9)

R code for Figure 2E

R code for figure 2D

library(ggalluvial)

library(ggplot2)

library(dplyr)

inputFile="Sanky plot input.txt"

outFile="ggalluvial.pdf"

setwd("D:\\ ")

rt=read.table(inputFile, header = T, sep="\t", check.names=F)

corLodes=to\_lodes\_form(rt, axes = 1:ncol(rt), id = "Cohort")

pdf(file=outFile,width=7,height=6)

mycol <- rep(c("#029149","#6E568C","#E0367A","#D8D155","#223D6C","#D20A13","#431A3D","#91612D","#FFD121","#088247","#11AA4D","#58CDD9","#7A142C","#5D90BA","#64495D","#7CC767"),15)

ggplot(corLodes, aes(x = x, stratum = stratum, alluvium = Cohort,fill = stratum, label = stratum)) +

scale\_x\_discrete(expand = c(0, 0)) +

geom\_flow(width = 2/10,aes.flow = "forward") +

geom\_stratum(alpha = .9,width = 2/10) +

scale\_fill\_manual(values = mycol) +

geom\_text(stat = "stratum", size = 2,color="black") +

xlab("") + ylab("") + theme\_bw() +

theme(axis.line = element\_blank(),axis.ticks = element\_blank(),axis.text.y = element\_blank())

theme(panel.grid =element\_blank()) +

theme(panel.border = element\_blank()) +

ggtitle("") + guides(fill = FALSE)

dev.off()

R code for figure 3A-B

library(survival)

library(survminer)

setwd("D:\\ ")

bioSurvival=function(inputFile=null,outFile=null){

rt=read.table(inputFile, header=T, sep="\t")

diff=survdiff(Surv(futime, fustat) ~risk,data = rt)

pValue=1-pchisq(diff$chisq,df=1)

if(pValue<0.001){

pValue="p<0.001"

}else{

pValue=paste0("p=",sprintf("%.03f",pValue))

}

fit <- survfit(Surv(futime, fustat) ~ risk, data = rt)

surPlot=ggsurvplot(fit,

data=rt,

conf.int=T,

pval=pValue,

pval.size=6,

legend.title="Risk",

legend.labs=c("High risk", "Low risk"),

xlab="Time(years)",

break.time.by = 1,

palette=c("red", "blue"),

risk.table=TRUE,

risk.table.title="",

risk.table.col = "strata",

risk.table.height=.25)

pdf(file=outFile,onefile = FALSE,width = 6.5,height =5.5)

print(surPlot)

dev.off()

}

bioSurvival(inputFile="trainRisk.txt", outFile="trainSurv.pdf")

bioSurvival(inputFile="testRisk.txt", outFile="testSurv.pdf")

R code for figure 3C-D

library(survival)

library(survminer)

library(timeROC)

riskFile="risk.all.txt" #风险文件

cliFile="clinical.txt" #临床数据文件

setwd("") #修改工作目录

#读取风险输入文件

risk=read.table(riskFile, header=T, sep="\t", check.names=F, row.names=1)

risk=risk[,c("futime", "fustat", "riskScore")]

#读取临床数据文件

cli=read.table(cliFile, header=T, sep="\t", check.names=F, row.names=1)

#合并数据

samSample=intersect(row.names(risk), row.names(cli))

risk1=risk[samSample,,drop=F]

cli=cli[samSample,,drop=F]

rt=cbind(risk1, cli)

#定义颜色

bioCol=rainbow(ncol(rt)-1, s=0.9, v=0.9)

######绘制1 3 5年的ROC曲线######

ROC\_rt=timeROC(T=risk$futime,delta=risk$fustat,

marker=risk$riskScore,cause=1,

weighting='aalen',

times=c(1,3,5),ROC=TRUE)

pdf(file="ROC.pdf", width=5, height=5)

plot(ROC\_rt,time=1,col=bioCol[1],title=FALSE,lwd=2)

plot(ROC\_rt,time=3,col=bioCol[2],add=TRUE,title=FALSE,lwd=2)

plot(ROC\_rt,time=5,col=bioCol[3],add=TRUE,title=FALSE,lwd=2)

legend('bottomright',

c(paste0('AUC at 1 years: ',sprintf("%.03f",ROC\_rt$AUC[1])),

paste0('AUC at 3 years: ',sprintf("%.03f",ROC\_rt$AUC[2])),

paste0('AUC at 5 years: ',sprintf("%.03f",ROC\_rt$AUC[3]))),

col=bioCol[1:3], lwd=2, bty = 'n')

dev.off()

######绘制临床的ROC曲线######

predictTime=1 #定义预测年限

aucText=c()

pdf(file="cliROC.pdf", width=6, height=6)

#绘制风险得分的ROC曲线

i=3

ROC\_rt=timeROC(T=risk$futime,

delta=risk$fustat,

marker=risk$riskScore, cause=1,

weighting='aalen',

times=c(predictTime),ROC=TRUE)

plot(ROC\_rt, time=predictTime, col=bioCol[i-2], title=FALSE, lwd=2)

aucText=c(paste0("Risk", ", AUC=", sprintf("%.3f",ROC\_rt$AUC[2])))

abline(0,1)

#对临床数据进行循环，绘制临床数据的ROC曲线

for(i in 4:ncol(rt)){

ROC\_rt=timeROC(T=rt$futime,

delta=rt$fustat,

marker=rt[,i], cause=1,

weighting='aalen',

times=c(predictTime),ROC=TRUE)

plot(ROC\_rt, time=predictTime, col=bioCol[i-2], title=FALSE, lwd=2, add=TRUE)

aucText=c(aucText, paste0(colnames(rt)[i],", AUC=",sprintf("%.3f",ROC\_rt$AUC[2])))

}

#绘制图例，得到ROC曲线下的面积

legend("bottomright", aucText,lwd=2,bty="n",col=bioCol[1:(ncol(rt)-1)])

dev.off()

R code for Figure 3E-F

library(survival) #引用包

setwd("") #设置工作目录

############绘制森林图函数############

bioForest=function(coxFile=null, forestFile=null, forestCol=null){

#读取输入文件

rt <- read.table(coxFile, header=T, sep="\t", check.names=F, row.names=1)

gene <- rownames(rt)

hr <- sprintf("%.3f",rt$"HR")

hrLow <- sprintf("%.3f",rt$"HR.95L")

hrHigh <- sprintf("%.3f",rt$"HR.95H")

Hazard.ratio <- paste0(hr,"(",hrLow,"-",hrHigh,")")

pVal <- ifelse(rt$pvalue<0.001, "<0.001", sprintf("%.3f", rt$pvalue))

#输出图形

pdf(file=forestFile, width=6.6, height=4.5)

n <- nrow(rt)

nRow <- n+1

ylim <- c(1,nRow)

layout(matrix(c(1,2),nc=2),width=c(3,2.5))

#绘制森林图左边的临床信息

xlim = c(0,3)

par(mar=c(4,2.5,2,1))

plot(1,xlim=xlim,ylim=ylim,type="n",axes=F,xlab="",ylab="")

text.cex=0.8

text(0,n:1,gene,adj=0,cex=text.cex)

text(1.5-0.5\*0.2,n:1,pVal,adj=1,cex=text.cex);text(1.5-0.5\*0.2,n+1,'pvalue',cex=text.cex,font=2,adj=1)

text(3.1,n:1,Hazard.ratio,adj=1,cex=text.cex);text(3.1,n+1,'Hazard ratio',cex=text.cex,font=2,adj=1)

#绘制右边的森林图

par(mar=c(4,1,2,1),mgp=c(2,0.5,0))

xlim = c(0,max(as.numeric(hrLow),as.numeric(hrHigh)))

plot(1,xlim=xlim,ylim=ylim,type="n",axes=F,ylab="",xaxs="i",xlab="Hazard ratio")

arrows(as.numeric(hrLow),n:1,as.numeric(hrHigh),n:1,angle=90,code=3,length=0.05,col="darkblue",lwd=2.5)

abline(v=1,col="black",lty=2,lwd=2)

boxcolor = ifelse(as.numeric(hr) > 1, forestCol, forestCol)

points(as.numeric(hr), n:1, pch = 15, col = boxcolor, cex=1.5)

axis(1)

dev.off()

}

############绘制森林图函数############

############独立预后分析函数#############

indep=function(riskFile=null, cliFile=null, project=null){

risk=read.table(riskFile, header=T, sep="\t", check.names=F, row.names=1) #读取风险文件

cli=read.table(cliFile, header=T, sep="\t", check.names=F, row.names=1) #读取临床文件

#数据合并

sameSample=intersect(row.names(cli),row.names(risk))

risk=risk[sameSample,]

cli=cli[sameSample,]

rt=cbind(futime=risk[,1], fustat=risk[,2], cli, riskScore=risk[,(ncol(risk)-1)])

#单因素独立预后分析

uniCoxFile=paste0(project,".uniCox.txt")

uniCoxPdf=paste0(project,".uniCox.pdf")

uniTab=data.frame()

for(i in colnames(rt[,3:ncol(rt)])){

cox <- coxph(Surv(futime, fustat) ~ rt[,i], data = rt)

coxSummary = summary(cox)

uniTab=rbind(uniTab,

cbind(id=i,

HR=coxSummary$conf.int[,"exp(coef)"],

HR.95L=coxSummary$conf.int[,"lower .95"],

HR.95H=coxSummary$conf.int[,"upper .95"],

pvalue=coxSummary$coefficients[,"Pr(>|z|)"])

)

}

write.table(uniTab,file=uniCoxFile,sep="\t",row.names=F,quote=F)

bioForest(coxFile=uniCoxFile, forestFile=uniCoxPdf, forestCol="green")

#多因素独立预后分析

multiCoxFile=paste0(project,".multiCox.txt")

multiCoxPdf=paste0(project,".multiCox.pdf")

uniTab=uniTab[as.numeric(uniTab[,"pvalue"])<1,]

rt1=rt[,c("futime","fustat",as.vector(uniTab[,"id"]))]

multiCox=coxph(Surv(futime, fustat) ~ ., data = rt1)

multiCoxSum=summary(multiCox)

multiTab=data.frame()

multiTab=cbind(

HR=multiCoxSum$conf.int[,"exp(coef)"],

HR.95L=multiCoxSum$conf.int[,"lower .95"],

HR.95H=multiCoxSum$conf.int[,"upper .95"],

pvalue=multiCoxSum$coefficients[,"Pr(>|z|)"])

multiTab=cbind(id=row.names(multiTab),multiTab)

write.table(multiTab, file=multiCoxFile, sep="\t", row.names=F, quote=F)

bioForest(coxFile=multiCoxFile, forestFile=multiCoxPdf, forestCol="red")

}

R code for Figure 4A-B

#install.packages("survival")

#install.packages("regplot")

#install.packages("rms")

#引用包

library(survival)

library(regplot)

library(rms)

riskFile="risk.all.txt" #风险输入文件

cliFile="clinical.txt" #临床数据文件

setwd("") #修改工作目录

#读取风险输入文件

risk=read.table(riskFile, header=T, sep="\t", check.names=F, row.names=1)

#读取临床数据文件

cli=read.table(cliFile, header=T, sep="\t", check.names=F, row.names=1)

cli=cli[apply(cli,1,function(x)any(is.na(match('unknow',x)))),,drop=F]

cli$Age=as.numeric(cli$Age)

#合并数据

samSample=intersect(row.names(risk), row.names(cli))

risk1=risk[samSample,,drop=F]

cli=cli[samSample,,drop=F]

rt=cbind(risk1[,c("futime", "fustat", "risk")], cli)

#绘制列线图

res.cox=coxph(Surv(futime, fustat) ~ . , data = rt)

nom1=regplot(res.cox,

plots = c("density", "boxes"),

clickable=F,

title="",

points=TRUE,

droplines=TRUE,

observation=rt[15,],

rank="sd",

failtime = c(1,3,5),

prfail = F)

#列线图风险打分

nomoRisk=predict(res.cox, data=rt, type="risk")

rt=cbind(risk1, Nomogram=nomoRisk)

outTab=rbind(ID=colnames(rt), rt)

write.table(outTab, file="nomoRisk.txt", sep="\t", col.names=F, quote=F)

#校准曲线

pdf(file="calibration.pdf", width=5, height=5)

#1年校准曲线

f <- cph(Surv(futime, fustat) ~ Nomogram, x=T, y=T, surv=T, data=rt, time.inc=1)

cal <- calibrate(f, cmethod="KM", method="boot", u=1, m=(nrow(rt)/3), B=1000)

plot(cal, xlim=c(0,1), ylim=c(0,1),

xlab="Nomogram-predicted OS (%)", ylab="Observed OS (%)", lwd=1.5, col="green", sub=F)

#3年校准曲线

f <- cph(Surv(futime, fustat) ~ Nomogram, x=T, y=T, surv=T, data=rt, time.inc=3)

cal <- calibrate(f, cmethod="KM", method="boot", u=3, m=(nrow(rt)/3), B=1000)

plot(cal, xlim=c(0,1), ylim=c(0,1), xlab="", ylab="", lwd=1.5, col="blue", sub=F, add=T)

#5年校准曲线

f <- cph(Surv(futime, fustat) ~ Nomogram, x=T, y=T, surv=T, data=rt, time.inc=5)

cal <- calibrate(f, cmethod="KM", method="boot", u=5, m=(nrow(rt)/3), B=1000)

plot(cal, xlim=c(0,1), ylim=c(0,1), xlab="", ylab="", lwd=1.5, col="red", sub=F, add=T)

legend('bottomright', c('1-year', '3-year', '5-year'),

col=c("green","blue","red"), lwd=1.5, bty = 'n')

dev.off()

R code for Figure 3G-I and Figure 4C-E

#install.packages("survival")

#install.packages("survminer")

#install.packages("timeROC")

#install.packages("ggDCA")

#引用包

library(survival)

library(survminer)

library(timeROC)

library(ggDCA)

predictTime=1 #预测时间

riskFile="nomoRisk.txt" #列线图风险输入文件

cliFile="clinical.txt" #临床数据文件

setwd("") #修改工作目录

#读取风险输入文件

risk=read.table(riskFile, header=T, sep="\t", check.names=F, row.names=1)

#读取临床数据文件

cli=read.table(cliFile, header=T, sep="\t", check.names=F, row.names=1)

#合并数据

samSample=intersect(row.names(risk), row.names(cli))

risk1=risk[samSample,,drop=F]

cli1=cli[samSample,,drop=F]

data=cbind(risk1, cli1)

#DCA分析

rt=cbind(risk1[,c("futime","fustat","Risk","Nomogram")], cli1)

rt[,"Age"]=ifelse(rt[,"Age"]>65, 1, 0)

rt[,"Nomogram"]=ifelse(rt[,"Nomogram"]>median(rt[,"Nomogram"]), 1, 0)

Nomogram<-coxph(Surv(futime,fustat)~Nomogram,rt)

Risk<-coxph(Surv(futime,fustat)~Risk,rt)

Age<-coxph(Surv(futime,fustat)~Age,rt)

Gender<-coxph(Surv(futime,fustat)~Gender,rt)

Grade<-coxph(Surv(futime,fustat)~Grade,rt)

Stage<-coxph(Surv(futime,fustat)~Stage,rt)

#绘制决策曲线

pdf(file="DCA.pdf", width=6.5, height=5.2)

d\_train=dca(Nomogram,Risk,Age,Gender,Grade,Stage, times=predictTime)

ggplot(d\_train, linetype=1)

dev.off()

######绘制临床的ROC曲线######

rt=cbind(risk1[,c("futime","fustat","riskScore","Nomogram")], cli1)

aucText=c()

bioCol=rainbow(ncol(rt)-1, s=0.9, v=0.9)

pdf(file="cliROC.pdf", width=6, height=6)

#绘制风险得分的ROC曲线

i=3

ROC\_rt=timeROC(T=risk$futime,

delta=risk$fustat,

marker=risk$riskScore, cause=1,

weighting='aalen',

times=c(predictTime),ROC=TRUE)

plot(ROC\_rt, time=predictTime, col=bioCol[i-2], title=FALSE, lwd=2)

aucText=c(paste0("Risk", ", AUC=", sprintf("%.3f",ROC\_rt$AUC[2])))

abline(0,1)

#对临床数据进行循环，绘制临床数据的ROC曲线

for(i in 4:ncol(rt)){

ROC\_rt=timeROC(T=rt$futime,

delta=rt$fustat,

marker=rt[,i], cause=1,

weighting='aalen',

times=c(predictTime),ROC=TRUE)

plot(ROC\_rt, time=predictTime, col=bioCol[i-2], title=FALSE, lwd=2, add=TRUE)

aucText=c(aucText, paste0(colnames(rt)[i],", AUC=",sprintf("%.3f",ROC\_rt$AUC[2])))

}

#绘制图例，得到ROC曲线下的面积

legend("bottomright", aucText,lwd=2,bty="n",col=bioCol[1:(ncol(rt)-1)])

dev.off()

R code for Figure 3J

#install.packages("survival")

#install.packages("rms")

#install.packages("pec")

#引用包

library(dplyr)

library(survival)

library(rms)

library(pec)

riskFile="risk.all.txt" #风险文件

cliFile="clinical.txt" #临床数据文件

setwd("") #修改工作目录

#读取风险输入文件

risk=read.table(riskFile, header=T, sep="\t", check.names=F, row.names=1)

risk=risk[,c("futime", "fustat", "riskScore")]

#读取临床数据文件

cli=read.table(cliFile, header=T, sep="\t", check.names=F, row.names=1)

#合并数据

samSample=intersect(row.names(risk), row.names(cli))

risk1=risk[samSample,,drop=F]

cli=cli[samSample,,drop=F]

rt=cbind(risk1, cli)

#定义颜色

bioCol=rainbow(ncol(rt)-1, s=0.9, v=0.9)

#C-index值计算

riskScore=cph(Surv(futime,fustat)~riskScore, data=rt, surv=TRUE)

Age=cph(Surv(futime,fustat)~Age, data=rt, surv=TRUE)

Gender=cph(Surv(futime,fustat)~Gender, data=rt, surv=TRUE)

Grade=cph(Surv(futime,fustat)~Grade, data=rt, surv=TRUE)

Stage=cph(Surv(futime,fustat)~Stage, data=rt, surv=TRUE)

c\_index <- cindex(list("Risk score"=riskScore,

"Age"=Age,

"Gender"=Gender,

"Grade"=Grade,

"Stage"=Stage),

formula=Surv(futime,fustat)~ .,

data=rt,

eval.times=seq(0,10,1),

splitMethod="bootcv",

B=1000

)

#输出图形

pdf(file="C-index.pdf", width=5.5, height=5)

plot(c\_index, xlim=c(0,10), ylim=c(0.4,0.8), col=bioCol, legend.x=6, legend.y=0.82, legend.cex=1)

dev.off()

R code for Figure 5

library(survminer)

riskFile="allRisk.txt"

cliFile="clinical.txt"

setwd("D:\\ ")

risk=read.table(riskFile, header=T, sep="\t", check.names=F, row.names=1)

cli=read.table(cliFile, header=T, sep="\t", check.names=F, row.names=1)

sameSample=intersect(row.names(cli), row.names(risk))

risk=risk[sameSample,]

cli=cli[sameSample,]

data=cbind(futime=risk[,1],fustat=risk[,2],cli,risk=risk[,"risk"])

for(i in colnames(data[,3:(ncol(data)-1)])){

rt=data[,c("futime","fustat",i,"risk")]

rt=rt[(rt[,i]!="unknow"),]

colnames(rt)=c("futime","fustat","clinical","risk")

tab=table(rt[,"clinical"])

tab=tab[tab!=0]

for(j in names(tab)){

rt1=rt[(rt[,"clinical"]==j),]

tab1=table(rt1[,"risk"])

tab1=tab1[tab1!=0]

labels=names(tab1)

if(length(labels)==2){

titleName=j

if((i=="age") | (i=="Age") | (i=="AGE")){

titleName=paste0("age",j)

}

diff=survdiff(Surv(futime, fustat) ~risk,data = rt1)

pValue=1-pchisq(diff$chisq,df=1)

if(pValue<0.001){

pValue="p<0.001"

}else{

pValue=paste0("p=",sprintf("%.03f",pValue))

}

fit <- survfit(Surv(futime, fustat) ~ risk, data = rt1)

surPlot=ggsurvplot(fit,

data=rt1,

conf.int=F,

pval=pValue,

pval.size=6,

title=paste0("Patients with ",titleName),

legend.title="Risk",

legend.labs=labels,

font.legend=12,

xlab="Time(years)",

break.time.by = 1,

palette=c("red", "blue"),

risk.table=TRUE,

risk.table.title="",

risk.table.col = "strata",

risk.table.height=.25)

j=gsub(">=","ge",j);j=gsub("<=","le",j);j=gsub(">","gt",j);j=gsub("<","lt",j)

pdf(file=paste0("survival.",i,"\_",j,".pdf"),onefile = FALSE,

width = 6,

height =5)

print(surPlot)

dev.off()

R code for Figure 7A-F

library(limma)

library(ggpubr)

library(pRRophetic)

library(ggplot2)

set.seed(12345)

expFile="symbol.txt"

riskFile="allRisk.txt"

drug="Rapamycin"

setwd("D:\\")

rt = read.table(expFile, header=T, sep="\t", check.names=F)

rt=as.matrix(rt)

rownames(rt)=rt[,1]

exp=rt[,2:ncol(rt)]

dimnames=list(rownames(exp),colnames(exp))

data=matrix(as.numeric(as.matrix(exp)),nrow=nrow(exp),dimnames=dimnames)

data=avereps(data)

data=data[rowMeans(data)>0.5,]

group=sapply(strsplit(colnames(data),"\\-"), "[", 4)

group=sapply(strsplit(group,""), "[", 1)

group=gsub("2","1",group)

data=data[,group==0]

data=t(data)

rownames(data)=gsub("(.\*?)\\-(.\*?)\\-(.\*?)\\-(.\*)", "\\1\\-\\2\\-\\3", rownames(data))

data=avereps(data)

data=t(data)

senstivity=pRRopheticPredict(data, drug, selection=1)

senstivity=senstivity[senstivity!="NaN"]

#senstivity[senstivity>quantile(senstivity,0.99)]=quantile(senstivity,0.99)

risk=read.table(riskFile, header=T, sep="\t", check.names=F, row.names=1)

sameSample=intersect(row.names(risk), names(senstivity))

risk=risk[sameSample, "risk",drop=F]

senstivity=senstivity[sameSample]

rt=cbind(risk, senstivity)

rt$risk=factor(rt$risk, levels=c("low", "high"))

type=levels(factor(rt[,"risk"]))

comp=combn(type, 2)

my\_comparisons=list()

for(i in 1:ncol(comp)){my\_comparisons[[i]]<-comp[,i]}

boxplot=ggboxplot(rt, x="risk", y="senstivity", fill="risk",

xlab="Risk",

ylab=paste0(drug, " senstivity (IC50)"),

legend.title="Risk",

palette=c("green", "red")

)+

stat\_compare\_means(comparisons=my\_comparisons)

pdf(file=paste0(drug, ".pdf"), width=5, height=4.5)

print(boxplot)

dev.off()

R code for figure 7

library(limma)

library(vioplot)

immuneFile="CIBERSORT-Results.txt"

cluFile="cluster.txt"

pFilter=0.05

setwd("D:\\ ")

immune=read.table(immuneFile, header=T, sep="\t", check.names=F, row.names=1)

immune=immune[immune[,"P-value"]<pFilter,]

immune=as.matrix(immune[,1:(ncol(immune)-3)])

group=sapply(strsplit(row.names(immune),"\\-"), "[", 4)

group=sapply(strsplit(group,""), "[", 1)

group=gsub("2", "1", group)

immune=immune[group==0,]

row.names(immune)=gsub("(.\*?)\\-(.\*?)\\-(.\*?)\\-(.\*?)\\-.\*", "\\1\\-\\2\\-\\3", row.names(immune))

immune=avereps(immune)

cluster=read.table(cluFile, header=F, sep="\t", row.names=1, check.names=F)

lowName=row.names(cluster)[cluster[,1]==1]

highName=row.names(cluster)[cluster[,1]==2]

lowImm=intersect(row.names(immune), lowName)

highImm=intersect(row.names(immune), highName)

rt=rbind(immune[lowImm,], immune[highImm,])

lowNum=length(lowImm)

highNum=length(highImm)

outTab=data.frame()

pdf("vioplot.pdf", width=13, height=9)

par(las=1,mar=c(10,6,3,3))

x=c(1:ncol(rt))

y=c(1:ncol(rt))

plot(x, y,

xlim=c(0,63), ylim=c(min(rt),max(rt)+0.02),

main="",xlab="", ylab="Fraction",

pch=21,

col="white",

xaxt="n")

bioCol=c("#0066FF","#FF0000","#FF9900","#6E568C","#7CC767","#223D6C","#D20A13","#FFD121","#088247","#11AA4D")

for(i in 1:ncol(rt)){

if(sd(rt[1:lowNum,i])==0){

rt[1,i]=0.00001

}

if(sd(rt[(lowNum+1):(lowNum+highNum),i])==0){

rt[(lowNum+1),i]=0.00001

}

lowData=rt[1:lowNum,i]

highData=rt[(lowNum+1):(lowNum+highNum),i]

vioplot(lowData,at=3\*(i-1),lty=1,add = T,col=bioCol[1])

vioplot(highData,at=3\*(i-1)+1,lty=1,add = T,col=bioCol[2])

wilcoxTest=wilcox.test(lowData,highData)

p=wilcoxTest$p.value

if(p<pFilter){

cellPvalue=cbind(Cell=colnames(rt)[i],pvalue=p)

outTab=rbind(outTab,cellPvalue)

}

mx=max(c(lowData,highData))

lines(c(x=3\*(i-1)+0.2,x=3\*(i-1)+0.8),c(mx,mx))

text(x=3\*(i-1)+0.5, y=mx+0.02, labels=ifelse(p<0.001, paste0("p<0.001"), paste0("p=",sprintf("%.03f",p))), cex = 0.8)

}

legend("topright",

c("Cluster1", "Cluster2"),

lwd=4.5,bty="n",cex=1.5,

col=bioCol[1:2])

text(seq(1,64,3),-0.05,xpd = NA,labels=colnames(rt),cex = 0.9,srt = 45,pos=2)

dev.off()

write.table(outTab,file="diff.result.txt",sep="\t",row.names=F,quote=F)

library(corrplot)

cellFile="CIBERSORT-Results.txt"

scoreFile="scores.txt"

setwd("D:\\ ")

cell=read.table(cellFile, header=T, sep="\t", check.names=F, row.names=1)

score=read.table(scoreFile, header=T, sep="\t", check.names=F, row.names=1)

sameSample=intersect(row.names(cell), row.names(score))

cell=cell[sameSample,]

score=score[sameSample,]

data=t(cbind(cell, score[,1:2]))

data=t(data)

M=cor(data)

res1=cor.mtest(data, conf.level = 0.95)

pdf(file="immCor.pdf", width=8, height=8)

corrplot(M,

order="original",

method = "pie",

type = "upper",

tl.cex=0.8, pch=T,

p.mat = res1$p,

insig = "blank",

sig.level=0.05,

col=colorRampPalette(c("blue", "white", "red"))(50),

tl.col="black")

dev.off()