################# Consensus map clustered via the NMF ################

setwd("D:\\ ")

1. ####### NFM #######

library(survival)

library(NMF)

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

pFilter=0.01

# Univariate Cox analysis

sigGenes=c()

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

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

coxSummary=summary(cox)

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

if(coxP<0.01){ sigGenes=c(sigGenes,i) }

}

#NMF analysis

data=t(rt[,sigGenes])

res=nmf(data, rank=2:10, method="brunet", nrun=10, seed=123456)

pdf(file="cophenetic.pdf", width=8, height=7, onefile=F)

plot(res)

dev.off()

# Output all clustering results

pdf(file="heatmap.all.pdf", width=15, height=15, onefile=F)

consensusmap(res,

annRow=NA,

annCol=NA,

tracks=c("consensus:"),

main="Consensus matrix",

info=FALSE)

dev.off()

# Output clustering results

clusterNum=2

res=nmf(data, rank=clusterNum, method="brunet", nrun=10, seed=123456)

Cluster=predict(res)

Cluster=as.data.frame(Cluster)

Cluster$Cluster=paste0("C", Cluster$Cluster)

clusterOut=rbind(ID=colnames(Cluster), Cluster)

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

# heatmap (Figure 1A)

pdf(file="heatmap.pdf", width=6, height=6, onefile=F)

consensusmap(res,

annRow=NA,

annCol=NA,

tracks=c("consensus:"),

main="Consensus matrix",

info=FALSE)

dev.off()

2. ######## Overall survival (Figure 1B) #######

library(survival)

library(survminer)

clusterFile="cluster.txt"

cliFile="time.txt"

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

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

colnames(cli)=c("futime", "fustat")

cli$futime=cli$futime/12

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

rt=cbind(cli[sameSample,,drop=F], cluster[sameSample,,drop=F])

length=length(levels(factor(rt$Cluster)))

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

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

if(pValue<0.001){

pValue="p<0.001"

}else{

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

}

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

#print(surv\_median(fit))

bioCol=c("#519DCC","#FE5B64")

bioCol=bioCol[1:length]

surPlot=ggsurvplot(fit,

data=rt,

conf.int=F,

pval=paste("Log-rank test",pValue),

pval.size=5,

pval.method = TRUE,

legend.title="Cluster",

legend.labs=levels(factor(rt[,"Cluster"])),

legend = c(0.75, 0.75),

font.title = c(18, "bold", "darkblue"),

font.x=c(13, "bold.italic", "red"),

font.y=c(13, "bold.italic", "darkred"),

font.legend=12,

font.tickslab=12,

risk.table.fontsize = 4,

risk.table.font.x=12,

risk.table.font.tickslab=12,

xlab="Time(years)",

ylab="Overall survival",

surv.median.line = "hv",

break.time.by = 1,

palette = bioCol,

risk.table=T,

tables.y.text = FALSE,

cumevents=F,

ggtheme = theme\_classic(),

risk.table.height=.25)

pdf(file="OS.pdf",onefile = FALSE,width=5,height=5)

print(surPlot)

dev.off()

3. ####### Progression free survival (Figure 1C) #######

library(survival)

library(survminer)

clusterFile="cluster.txt"

cliFile="Survival\_SupplementalTable\_S1\_20171025\_xena\_sp"

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

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

cli=cli[,c("PFI.time", "PFI")]

colnames(cli)=c("futime", "fustat")

cli$futime=cli$futime/365

cli=as.matrix(cli)

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

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

rt=cbind(cli[sameSample,,drop=F], cluster[sameSample,,drop=F])

length=length(levels(factor(rt$Cluster)))

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

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

if(pValue<0.001){

pValue="p<0.001"

}else{

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

}

fit=survfit(Surv(futime, fustat) ~ Cluster, data = rt)

#print(surv\_median(fit))

bioCol=c("#519DCC","#FE5B64")

bioCol=bioCol[1:length]

surPlot=ggsurvplot(fit,

data=rt,

conf.int=F,

pval=paste("Log-rank test",pValue),

pval.size=5,

pval.method = TRUE,

legend.title="Cluster",

legend.labs=levels(factor(rt[,"Cluster"])),

legend = c(0.75, 0.75),

font.title = c(18, "bold", "darkblue"),

font.x=c(13, "bold.italic", "red"),

font.y=c(13, "bold.italic", "darkred"),

font.legend=12,

font.tickslab=12,

risk.table.fontsize = 4,

risk.table.font.x=12,

risk.table.font.tickslab=12,

xlim=c(0,14),

xlab="Time(years)",

ylab="Progression free survival",

surv.median.line = "hv",

break.time.by = 1,

palette = bioCol,

risk.table=T,

tables.y.text = FALSE,

cumevents=F,

ggtheme = theme\_classic(),

risk.table.height=.25)

pdf(file="PFS.pdf",onefile=FALSE, width=5, height=5)

print(surPlot)

dev.off()

4. ####### Immune subtype distribution(Figure 1D) #######

library(ggalluvial)

library(ggplot2)

library(dplyr)

clusterFile="cluster.txt"

subtypeFile="Subtype\_Immune\_Model\_Based.txt"

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

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

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

sameSample=intersect(row.names(subtype), row.names(cluster))

subtype=subtype[sameSample,,drop=F]

subtype[,1]=gsub(".+\\((.+?)\\)", "\\1", subtype[,1])

colnames(subtype)=c("Immune subtype")

cluster=cluster[sameSample,,drop=F]

data=cbind(cluster, subtype)

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

pdf(file="ggalluvial.pdf", width=6, height=7)

mycol=rep(c("#519DCC","#FE5B64","#A396D4","#EA921D","#D14039","#029149","#6E568C","#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)) +

# aes.flow

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

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

scale\_fill\_manual(values = mycol) +

#size=3

geom\_text(stat = "stratum", size = 4,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()

5. ####### Cibersort: estimation of immune cells infiltration (Figure 1E) #######

library(reshape2)

library(ggpubr)

library(limma)

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

ssgseaScore=read.table("CIBERSORT-Results.txt", header=T, sep="\t", check.names=F, row.names=1)

sameSample=intersect(row.names(ssgseaScore), row.names(cluster))

ssgseaScore=ssgseaScore[sameSample,,drop=F]

cluster=cluster[sameSample,,drop=F]

scoreCluster=cbind(ssgseaScore, cluster)

data=melt(scoreCluster, id.vars=c("Cluster"))

colnames(data)=c("Cluster", "Immune", "Fraction")

bioCol=c("#519DCC","#FE5B64")

bioCol=bioCol[1:length(levels(factor(data[,"Cluster"])))]

p=ggboxplot(data, x="Immune", y="Fraction", fill="Cluster",

ylab="Immune infiltration (%)",

xlab="",

legend.title="Cluster",

palette=bioCol,

width=0.65,

outlier.size=0.5)+

theme\_classic()+

theme(legend.title = element\_text(size = rel(1.0)),

legend.position = "top",

legend.box = "vertical",

legend.text = element\_text(size = rel(1.0)))+

theme(axis.title.y = element\_text(size = rel(1.0)),

axis.text.y = element\_text(size = rel(1.0)),

axis.text.x = element\_text(size = rel(1.0)))

p=p+rotate\_x\_text(50)

pdf(file="Cibersort.pdf", width=8, height=4)

p+stat\_compare\_means(aes(group=Cluster),symnum.args=list(cutpoints = c(0, 0.001, 0.01, 0.05, 1), symbols = c("\*\*\*", "\*\*", "\*", "ns")),label = "p.signif")

dev.off()

################## Establishment of TME risk model #####################

6. ######Establishment of the model #######

library(survival)

library(caret)

library(glmnet)

library(survminer)

library(timeROC)

coxPfilter=0.05

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

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

#Univariate Cox analysis

outUniTab=data.frame()

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

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

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

coxSummary = summary(cox)

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

if(coxP<coxPfilter){

sigGenes=c(sigGenes,i)

outUniTab=rbind(outUniTab,

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|)"])

)

}

}

uniSigExp=train[,sigGenes]

uniSigExpOut=cbind(id=row.names(uniSigExp),uniSigExp)

write.table(outUniTab,file="uni.trainCox.txt",sep="\t",row.names=F,quote=F)

write.table(uniSigExpOut,file="uni.SigExp.txt",sep="\t",row.names=F,quote=F)

## LASSO regression analysis (Figure 2A, B)

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

y=data.matrix(Surv(uniSigExp$futime,uniSigExp$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]

lassoSigExp=uniSigExp[,c("futime", "fustat", lassoGene)]

lassoSigExpOut=cbind(id=row.names(lassoSigExp), lassoSigExp)

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

write.table(lassoSigExpOut,file="lasso.SigExp.txt",sep="\t",row.names=F,quote=F)

pdf("lasso.lambda.pdf")

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

dev.off()

pdf("lasso.cvfit.pdf")

plot(cvfit)

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

dev.off()

#############Construction of TME risk model#############

multiCox <- coxph(Surv(futime, fustat) ~ ., data = lassoSigExp)

multiCox=step(multiCox,direction = "both")

multiCoxSum=summary(multiCox)

outMultiTab=data.frame()

outMultiTab=cbind(

coef=multiCoxSum$coefficients[,"coef"],

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|)"])

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

outMultiTab=outMultiTab[,1:2]

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

#Output RS of TCGA train cohort

riskScore=predict(multiCox,type="risk",newdata=train) ##Using the train model to predict the RS of train cohort

coxGene=rownames(multiCoxSum$coefficients)

coxGene=gsub("`","",coxGene)

outCol=c("futime","fustat",coxGene)

medianTrainRisk=median(riskScore)

risk=as.vector(ifelse(riskScore>medianTrainRisk,"high","low"))

trainRiskOut=cbind(id=rownames(cbind(train[,outCol],riskScore,risk)),cbind(train[,outCol],riskScore,Risk=risk))

write.table(trainRiskOut,file="risk.TCGAtrain.txt",sep="\t",quote=F,row.names=F)

#Output RS of TCGA test cohort

riskScoreTest=predict(multiCox,type="risk",newdata=test) ##Using the train model to predict the RS of test cohort

riskTest=as.vector(ifelse(riskScoreTest>medianTrainRisk,"high","low"))

testRiskOut=cbind(id=rownames(cbind(test[,outCol],riskScoreTest,riskTest)),cbind(test[,outCol],riskScore=riskScoreTest,Risk=riskTest))

write.table(testRiskOut,file="risk.TCGAtest.txt",sep="\t",quote=F,row.names=F)

#Output RS of GSE13507

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

GEO$futime=GEO$futime/12

geoScore=predict(multiCox, type="risk", newdata=GEO)

geoRisk=as.vector(ifelse(geoScore>medianTrainRisk, "high", "low"))

GEO=cbind(GEO[,outCol], riskScore=as.vector(geoScore), Risk=geoRisk)

geoRiskOut=cbind(id=rownames(GEO), GEO)

write.table(geoRiskOut,file="risk.GEO.txt",sep="\t",quote=F,row.names=F)

#Output RS of the whole TCGA samples.

allRiskOut=rbind(trainRiskOut, testRiskOut)

write.table(allRiskOut,file="risk.TCGAall.txt",sep="\t",quote=F,row.names=F)

7. ###### Survival curve (Figure 2C, E, G and I) ######

library(survival)

library(survminer)

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

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

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=paste("Log-rank test",pValue),

pval.size=5,

legend.title=" ",

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

xlab="Time(years)",

#legend = c(0.8, 0.9),

font.title = c(15, "bold", "darkblue"),

font.x=c(13, "bold.italic", "#FF4040"),

font.y=c(13, "bold.italic", "darkred"),

font.legend=10,

font.tickslab=12,

risk.table.fontsize = 4,

risk.table.font.x=12,

risk.table.font.tickslab=10,

break.time.by = 1,

palette=c("#FB8072","#80B1D3"),

risk.table=TRUE,

tables.y.text = FALSE,

cumevents=F,

ggtheme = theme\_classic()+

theme(plot.title = element\_text(size=12,hjust=0.5)),

risk.table.col = "strata",

risk.table.height=.25)

pdf(file=outFile,onefile = FALSE,width = 5,height = 5)

print(surPlot)

dev.off()

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

print(surPlot)

dev.off()

}

bioSurvival(inputFile="risk.TCGAtrain.txt", outFile="surv.TCGAtrain.pdf")

bioSurvival(inputFile="risk.TCGAtest.txt", outFile="surv.TCGAtest.pdf")

bioSurvival(inputFile="risk.TCGAall.txt", outFile="surv.TCGAall.pdf")

bioSurvival(inputFile="risk.GEO.txt", outFile="surv.GEO.pdf")

8. ###### ROC curve (Figure 2D, F, H and J) ######

library(survival)

library(survminer)

library(timeROC)

bioROC=function(inputFile=null, rocFile=null){

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

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

marker=rt$riskScore,cause=1,

weighting='aalen',

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

pdf(file=rocFile,width=5,height=5)

plot(ROC\_rt,time=1,col='#4DBBD5',title=FALSE,lwd=2)

plot(ROC\_rt,time=3,col='#E64B35',add=TRUE,title=FALSE,lwd=2)

plot(ROC\_rt,time=5,col='#00A087',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=c("#4DBBD5",'#E64B35','#00A087'),lwd=2,bty = 'n')

dev.off()

}

bioROC(inputFile="risk.TCGAtrain.txt", rocFile="ROC.TCGAtrain.pdf")

bioROC(inputFile="risk.TCGAtest.txt", rocFile="ROC.TCGAtest.pdf")

bioROC(inputFile="risk.TCGAall.txt", rocFile="ROC.TCGAall.pdf")

bioROC(inputFile="risk.GEO.txt", rocFile="ROC.GEO.pdf")

9. ###### Forest plots(Figure 3A and B) ######

library(survival)

Indep=function(riskFile=null, cliFile=null, uniOutFile=null, multiOutFile=null){

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,]

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

#Unicox

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=uniOutFile,sep="\t",row.names=F,quote=F)

#Multicox

uniTab=uniTab[as.numeric(uniTab[,"pvalue"])<0.05,]

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=multiOutFile,sep="\t",row.names=F,quote=F)

}

Indep(riskFile="risk.TCGAall.txt",

cliFile="clinical.txt",

uniOutFile="uniCox.xls",

multiOutFile="multiCox.xls")

#forest plot

library("forestplot")

data <- read.table("uniCox.txt", header=T,sep="\t")

tabletext <- cbind(c("Unitivariate analysis",data$Variable),

c("pvalue",ifelse(data$pvalue<0.01,paste("<0.001"),sprintf("%.03f",data$pvalue))),

c("Hazard Ratio(95% CI)",

paste(format(data$HR,nsmall=2)," (",

format(data$HR.95L,nsmall = 2)," to ",

format(data$HR.95H,nsmall = 2),")",sep="")))

forestplot(labeltext=tabletext,mean=c(1,data$HR),

lower=c(1,data$HR.95L,NA),upper=c(1,data$HR.95H))

#Output files

pdf("forestplot.pdf",width=6,height = 4)

forestplot(labeltext=tabletext,

mean=c(1,data$HR),#HR

lower=c(1,data$HR.95L),

upper=c(1,data$HR.95H),

#title="Hazard Ratio",

graph.pos=2,

graphwidth = unit(.3,"npc"),

fn.ci\_norm="fpDrawDiamondCI",

col=fpColors(box="steelblue", lines="black", zero = "black"),

boxsize = 0.35,

lwd.ci=2,ci.vertices.height = 0.1,ci.vertices=TRUE,

zero=1,

lwd.zero=1,

xticks = c(0, 1, 2, 3, 4, 5, 6),

lwd.xaxis=2,

xlab=" <-Favors Benificial Favors Harmful-> ",

hrzl\_lines=list("2" = gpar(lwd=2, col="black"),

#"4" = gpar(lwd=60,lineend="butt", columns=c(1:4), col="#99999922"),

"11" = gpar(lwd=2, col="black")),

txt\_gp=fpTxtGp(label=gpar(cex=0.7),

ticks=gpar(cex=0.8),

xlab=gpar(cex = 0.65),

title=gpar(cex = 0.6)),

is.summary = c(T,rep(F,27)),

lineheight = unit(.85,"cm"),

colgap = unit(0.3,"cm"),

mar=unit(rep(1, times = 4), "cm"),

new\_page = F

)

dev.off()

10. ###### Nomogram (Figure 3C) ######

library(survival)

library(regplot)

library(rms)

riskFile="risk.TCGAall.txt"

cliFile="clinical2.txt"

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)

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

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

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)

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[1,],

rank="sd",

failtime = c(1,3,5),

prfail = F)

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)

#Calibration curve (Figure 3D)

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

#1-year

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="#4DBBD5", sub=F)

#3-year

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="#E64B35", sub=F, add=T)

#5-year

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="#00A087", sub=F, add=T)

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

col=c("#4DBBD5","#E64B35","#00A087"), lwd=1.5, bty = 'n')

dev.off()

11. ###### DCA analysis and ROC curve ######

# DCAanalysis (Figure 3E)

library(survival)

library(survminer)

library(timeROC)

library(ggDCA)

predictTime=1

riskFile="nomoRisk.txt"

cliFile="clinical3.txt"

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)

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

risk1=risk[samSample,,drop=F]

cli1=cli[samSample,,drop=F]

data=cbind(risk1, cli1)

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)

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

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

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)

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)+

labs(title=" DCA curve ")+

theme\_bw()+

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

theme(plot.title = element\_text(size = 16,hjust = 0.5),

axis.title = element\_text(size = 12))+

theme(plot.margin=unit(rep(1,4),'lines'))

dev.off()

# Clinical ROC curve (Figure 3F)

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)

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)

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])))

}

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

dev.off()

############# Comparison of different BLCA gene signatures ###############

12. ####### Survival curve and ROC #######

library(limma)

library(survival)

library(survminer)

library(timeROC)

expFile="symbol.txt"

riskFile="risk.TCGAall.txt"

geneFiles=c("Cao signature.txt", "Yin signature.txt", "Zhang signature.txt", "Chen signature.txt")

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,]

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

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

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

data=data[,group==0]

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

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

colnames(riskRT)=c("futime","fustat","TME signature")

for(i in geneFiles){

header=unlist(strsplit(i, "\\."))

gene=read.table(i, header=F, sep="\t", check.names=F)

sameGene=intersect(as.vector(gene[,1]), row.names(data))

data1=data[sameGene,]

colnames(data1)=gsub("(.\*?)\\-(.\*?)\\-(.\*?)\\-.\*", "\\1\\-\\2\\-\\3", colnames(data1))

data1=t(data1)

data1=avereps(data1)

cli=riskRT[,c("futime", "fustat")]

sameSample=intersect(row.names(data1), row.names(cli))

data1=data1[sameSample,]

cli=cli[sameSample,]

data1=cbind(cli,data1)

multiCox=coxph(Surv(futime, fustat) ~ ., data = data1)

riskScore=predict(multiCox,type="risk", newdata=data1)

data1=cbind(data1, riskScore)

data1=data1[row.names(riskRT),]

riskRT=cbind(riskRT, data1[,"riskScore"])

colnames(riskRT)[ncol(riskRT)]=header[[1]]

}

riskOut=rbind(ID=colnames(riskRT), riskRT)

write.table(riskOut, file="risk.models.txt", sep="\t", col.names=F, quote=F)

# Survival curve (Figure 4A-E)

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

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

rt$Type=ifelse(rt[,varName]>median(rt[,varName]), "high", "low")

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

surPlot=ggsurvplot(fit,

data=rt,

conf.int=F,

pval=paste("Log-rank test",pValue),

pval.size=5,

title=varName,

legend.title="Risk",

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

legend = c(0.8, 0.85),

xlab="Time(years)",

font.title = c(15, "bold", "darkblue"),

font.x=c(13, "bold.italic", "#FF4040"),

font.y=c(13, "bold.italic", "darkred"),

font.legend=12,

font.tickslab=12,

break.time.by = 1,

palette=c("#FB8072","#80B1D3"),

risk.table=F,

risk.table.title="",

risk.table.col = "strata",

risk.table.height=.25)

pdf(file=outFile, onefile = FALSE, width=4, height=4)

print(surPlot)

dev.off()

}

# ROC curve (Figure 4A-E)

bioROC=function(inputFile=null, outFile=null, varName=null){

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

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

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

weighting='aalen',

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

pdf(file=outFile, width=5, height=5)

plot(ROC\_rt,time=1,col='#4DBBD5',title=FALSE,lwd=2)

plot(ROC\_rt,time=3,col='#E64B35',add=TRUE,title=FALSE,lwd=2)

plot(ROC\_rt,time=5,col='#00A087',add=TRUE,title=FALSE,lwd=2)

text(0.75, 0.24, varName, cex=1.2)

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=c("#4DBBD5",'#E64B35','#00A087'),lwd=2,bty = 'n')

dev.off()

}

for(varName in colnames(riskRT)[3:ncol(riskRT)]){

bioSurvival(inputFile="risk.models.txt", outFile=paste0("sur.",varName,".pdf"), varName=varName)

bioROC(inputFile="risk.models.txt", outFile=paste0("ROC.",varName,".pdf"), varName=varName)

}

13.########## C-index and RMS curve ##########

# C-index (Figure 4F)

library(survival)

library(survcomp)

library(ggplot2)

library(ggpubr)

inputFile="risk.models.txt"

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

df=data.frame()

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

cindex=concordance.index(x=rt[,i], surv.time=rt$futime, surv.event=rt$fustat,method="noether")

df=rbind(df, cbind(i,sprintf("%.03f",cindex$c.index)))

}

colnames(df)=c("signature", "cindex")

df[,"cindex"]=as.numeric(df[,"cindex"])

#color=rainbow(nrow(df),alpha=0.75)

color=c("#E64B35","#EFC000","#00A087","#7AA6DC","#F39B7F")

p=ggbarplot(df, x="signature", y="cindex", fill="signature",

xlab="", ylab="C-index", add = "none",

palette=color,

label=T, legend="")

p=p+rotate\_x\_text(40)

p=p+ylim(0,round(max(df[,"cindex"])+0.15,1))

pdf(file="C-index.pdf", width=6, height=5)

print(p)

dev.off()

# RMS curve (Figure 4G)

outdata=list()

legendsname=c()

for(i in 3:ncol(rt)){

OS=Surv(rt$futime, rt$fustat)

marker=rt[,i]

marker.pp<-seq(from=0, to=1, length=100)

marker.qq<-quantile(marker,marker.pp)

fitdat.df<-data.frame(marker=marker)

newdat.df<-data.frame(marker=marker.qq)

cox.model<-coxph(OS~marker, data=fitdat.df)

rms.calc <-summary(survfit(cox.model, newdata=newdat.df))

rms.mean <-rms.calc$table[,"rmean"]

name=colnames(rt)[i]

HR=sprintf("%.03f", summary(cox.model)$conf.int[,"exp(coef)"])

HR.95L=sprintf("%.03f", summary(cox.model)$conf.int[,"lower .95"])

HR.95H=sprintf("%.03f", summary(cox.model)$conf.int[,"upper .95"])

pvalue=summary(cox.model)$coefficients[,"Pr(>|z|)"]

p=ifelse(pvalue<0.001,"p<0.001",paste0("p=",sprintf("%.03f",pvalue)))

legendsname=c(legendsname,paste0(name,", HR:",HR,"(",HR.95L,"-",HR.95H,"), ",p))

outdata[[name]]= data.frame(marker.pp,rms.mean)

}

alldata=do.call("rbind",outdata)

xlim2=max(alldata$rms.mean)

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

par(las=1)

plot(1,xlim=c(0,1),ylim=c(0,xlim2),type="n",xlab="Percentile of scores",ylab="RMS (years)")

names=names(outdata)

for(i in 1:length(outdata)){

namei=names[i]

outdatai=outdata[[namei]]

points(outdatai$marker.pp,outdatai$rms.mean,col=color[i],pch=20,cex=0.8)

}

legend("bottomleft",legend=legendsname,col=color,pch=20,bty="n",cex=1)

dev.off()

14. ########## GESA analysis (Figure 5A and B) ##########

library(limma)

library(org.Hs.eg.db)

library(clusterProfiler)

library(enrichplot)

expFile="symbol.txt"

riskFile="risk.TCGAall.txt"

gmtFile="c2.cp.kegg.v7.4.symbols.gmt"

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=t(avereps(data))

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

data=data[,row.names(Risk)]

dataL=data[,row.names(Risk[Risk[,"Risk"]=="low",])]

dataH=data[,row.names(Risk[Risk[,"Risk"]=="high",])]

meanL=rowMeans(dataL)

meanH=rowMeans(dataH)

meanL[meanL<0.00001]=0.00001

meanH[meanH<0.00001]=0.00001

logFC=log2(meanH)-log2(meanL)

logFC=sort(logFC,decreasing=T)

genes=names(logFC)

#Read geneset file

gmt=read.gmt(gmtFile)

# Enrichment analysis

kk=GSEA(logFC, TERM2GENE=gmt, pvalueCutoff = 1)

kkTab=as.data.frame(kk)

kkTab=kkTab[kkTab$pvalue<0.05,]

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

#High-risk enriched

termNum=5 #top 5 pathways

kkUp=kkTab[kkTab$NES>0,]

if(nrow(kkUp)>=termNum){

showTerm=row.names(kkUp)[1:termNum]

gseaplot=gseaplot2(kk, showTerm, base\_size=8, title="Enriched in high risk group")

pdf(file="GSEA.highRisk.pdf", width=7, height=5.5)

print(gseaplot)

dev.off()

}

#Low-risk enriched

termNum=5 #top 5 pathways

kkDown=kkTab[kkTab$NES<0,]

if(nrow(kkDown)>=termNum){

showTerm=row.names(kkDown)[1:termNum]

gseaplot=gseaplot2(kk, showTerm, base\_size=8, title="Enriched in low risk group")

pdf(file="GSEA.lowRisk.pdf", width=7, height=5.5)

print(gseaplot)

dev.off()

}

################## Immune Correlation analysis ##################

15.####### Boxplot (Figure 6A) ########

riskFile="risk.TCGAall.txt"

immFile=" CIBERSORT-Results.txt"

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

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

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

risk=as.data.frame(risk[sameSample,"Risk",drop=F])

immune=immune[sameSample,]

data=cbind(as.data.frame(immune), as.data.frame(risk))

data=melt(data, id.vars=c("Risk"))

colnames(data)=c("Risk", "Cell", "Infiltration")

boxplot=ggboxplot(data, x="Cell ", y=" Infiltration", fill = "Risk",

ylab="Immune infiltration (%)",

xlab="",

legend.title="Risk",

palette = c("#FB8072","#80B1D3"),

width=0.65,

outlier.size=0.5)+

theme\_classic()+

theme(legend.title = element\_text(size = rel(1.2)),

legend.position = "top",

legend.box = "vertical",

legend.text = element\_text(size = rel(1.2)))+

theme(axis.title.y = element\_text(size = rel(1.3))

)

p=p+rotate\_x\_text(60)

boxplot =p+stat\_compare\_means(aes(group=group),

symnum.args=list(cutpoints = c(0, 0.001, 0.01, 0.05, 1), symbols = c("\*\*\*", "\*\*", "\*", "ns")),

label = "p.signif")

pdf(file="boxplot.pdf", width=8, height=5)

print(boxplot)

dev.off()

16. ####### Corrlation plot (Figure 6B) ########

library(corrplot)

library(Hmisc)

riskFile="risk.TCGAall.txt"

immFile=" CIBERSORT-Results.txt"

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

risk=risk[,ncol(risk)-1,drop=F]

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

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

data=cbind(risk[sameSample,,drop=F], immune[sameSample,,drop=F])

res <- rcorr(as.matrix(data),type = "pearson")

N=res$r

M=res$P

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

corrplot(N,

order="original",

method = "square",

type = "lower",

tl.cex=0.8, pch=T,

tl.srt =45,

p.mat = M,

insig = "label\_sig",

pch.cex = 1.6,

sig.level = c(0.01,0.05),

number.cex = 1,

col=colorRampPalette(c("#084594","#9ECAE1","white","#FC9272","#99000D"))(50),

tl.col="black")

dev.off()

17. ######## Circos Plot (Figure 6C) #######

library(corrplot)

library(circlize)

tmbFile="TMB.txt"

riskFile="risk.TCGAall.txt"

immuneFile="CIBERSORT-Results.txt"

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

#data=t(data)

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

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

data=cbind(risk[sameSample,"riskScore",drop=F], data[sameSample,,drop=F])

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

sameSample=intersect(row.names(TMB), row.names(data))

rt=cbind(TMB[sameSample,,drop=F], data[sameSample,,drop=F])

cor1=cor(rt)

col = c(rgb(1,0,0,seq(1,0,length=32)),rgb(0,1,0,seq(0,1,length=32)))

cor1[cor1==1]=0

c1 = ifelse(c(cor1)>=0,rgb(1,0,0,abs(cor1)),rgb(0,1,0,abs(cor1)))

col1 = matrix(c1,nc=ncol(rt))

# Circos

pdf(file="circos.pdf", width=7, height=7)

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

circos.par(gap.degree=c(3,rep(2, nrow(cor1)-1)), start.degree = 180)

chordDiagram(cor1, grid.col=rainbow(ncol(rt)), col=col1, transparency = 0.5, symmetric = T)

par(xpd=T)

colorlegend(col, vertical = T,labels=c(1,0,-1),xlim=c(1.1,1.3),ylim=c(-0.4,0.4))

dev.off()

circos.clear()

18. ####### Boxplot of immunosuppressive genes (Figure 6D) #######

library(limma)

library(reshape2)

library(ggplot2)

library(ggpubr)

expFile="symbol.txt"

riskFile="risk.TCGAall.txt"

geneFile="Immunosuppressive.txt"

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)

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

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

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

data=data[,group==0]

geneRT=read.table(geneFile, header=F, sep="\t", check.names=F)

data=data[as.vector(geneRT[,1]),]

data=t(data)

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

data=avereps(data)

data=log2(data+1)

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

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

data=cbind(risk[sameSample,"Risk",drop=F], data[sameSample,,drop=F])

rt1=melt(data,id.vars=c("Risk"))

colnames(rt1)=c("Risk","Gene","Expression")

boxplot=ggboxplot(rt1, x="Gene", y="Expression", fill="Risk",

xlab="",

ylab="Relative expression",

legend.title="Risk",

width=0.65,

outlier.size=0.5,

palette = c("#FB8072","#80B1D3"))+

theme\_classic()+

theme(legend.title = element\_text(size = rel(1.2)),

legend.position = "top",

legend.box = "vertical",

legend.text = element\_text(size = rel(1.2)))+

theme(axis.title.y = element\_text(size = rel(1.2)),

axis.text.y = element\_text(size = rel(1.5)),

axis.text.x = element\_text(size = rel(1.5))

)+

rotate\_x\_text(50)+

stat\_compare\_means(aes(group=Risk),

method="kruskal.test",

symnum.args=list(cutpoints=c(0, 0.001, 0.01, 0.05, 1), symbols=c("\*\*\*", "\*\*", "\*", "ns")), label="p.signif")

pdf(file="Immunosuppressive.pdf", width=8, height=4)

print(boxplot)

dev.off()

19. ####### Boxplot of immune activation genes (Figure 6E) #######

library(limma)

library(reshape2)

library(ggplot2)

library(ggpubr)

expFile="symbol.txt"

riskFile="risk.TCGAall.txt"

geneFile="Immuneactivation.txt"

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)

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

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

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

data=data[,group==0]

geneRT=read.table(geneFile, header=F, sep="\t", check.names=F)

data=data[as.vector(geneRT[,1]),]

data=t(data)

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

data=avereps(data)

data=log2(data+1)

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

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

data=cbind(risk[sameSample,"Risk",drop=F], data[sameSample,,drop=F])

rt1=melt(data,id.vars=c("Risk"))

colnames(rt1)=c("Risk","Gene","Expression")

boxplot=ggboxplot(rt1, x="Gene", y="Expression", fill="Risk",

xlab="",

ylab="Relative expression",

legend.title="Risk",

width=0.65,

outlier.size=0.5,

palette = c("#FB8072","#80B1D3"))+

theme\_classic()+

theme(legend.title = element\_text(size = rel(1.2)),

legend.position = "top",

legend.box = "vertical",

legend.text = element\_text(size = rel(1.2)))+

theme(axis.title.y = element\_text(size = rel(1.2)),

axis.text.y = element\_text(size = rel(1.5)),

axis.text.x = element\_text(size = rel(1.5))

)+

rotate\_x\_text(50)+

stat\_compare\_means(aes(group=Risk),

method="kruskal.test",

symnum.args=list(cutpoints=c(0, 0.001, 0.01, 0.05, 1), symbols=c("\*\*\*", "\*\*", "\*", "ns")), label="p.signif")

pdf(file="Immuneactivation.pdf", width=12, height=4)

print(boxplot)

dev.off()

20. ####### Boxplot of MHC genes (Figure 6F) #######

library(limma)

library(reshape2)

library(ggplot2)

library(ggpubr)

expFile="symbol.txt"

riskFile="risk.TCGAall.txt"

geneFile="MHC.txt"

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)

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

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

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

data=data[,group==0]

geneRT=read.table(geneFile, header=F, sep="\t", check.names=F)

data=data[as.vector(geneRT[,1]),]

data=t(data)

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

data=avereps(data)

data=log2(data+1)

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

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

data=cbind(risk[sameSample,"Risk",drop=F], data[sameSample,,drop=F])

rt1=melt(data,id.vars=c("Risk"))

colnames(rt1)=c("Risk","Gene","Expression")

boxplot=ggboxplot(rt1, x="Gene", y="Expression", fill="Risk",

xlab="",

ylab="Relative expression",

legend.title="Risk",

width=0.65,

outlier.size=0.5,

palette = c("#FB8072","#80B1D3"))+

theme\_classic()+

theme(legend.title = element\_text(size = rel(1.2)),

legend.position = "top",

legend.box = "vertical",

legend.text = element\_text(size = rel(1.2)))+

theme(axis.title.y = element\_text(size = rel(1.2)),

axis.text.y = element\_text(size = rel(1.5)),

axis.text.x = element\_text(size = rel(1.5))

)+

rotate\_x\_text(50)+

stat\_compare\_means(aes(group=Risk),

method="kruskal.test",

symnum.args=list(cutpoints=c(0, 0.001, 0.01, 0.05, 1), symbols=c("\*\*\*", "\*\*", "\*", "ns")), label="p.signif")

pdf(file="MHC.pdf", width=6, height=4)

print(boxplot)

dev.off()

21. ####### Boxplot of IPS (Figure 7A) #######

library(limma)

library(ggplot2)

library("ggsignif")

library(reshape2)

library(ggpubr)

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

source("https://gist.githubusercontent.com/benmarwick/2a1bb0133ff568cbe28d/raw/fb53bd97121f7f9ce947837ef1a4c65a73bffb3f/geom\_flat\_violin.R")

my\_comparisons <- list(c("high","low"))

p=ggplot(rt, aes(x=Risk, y=IPS)) +

geom\_flat\_violin(aes(fill=Risk),position=position\_nudge(x=.25),color="black",adjust=2) +

geom\_jitter(aes(color=Risk), width=0.1,size = 0.5) +

geom\_boxplot(width=.1,position=position\_nudge(x=0.25),fill="white",size=0.5)+

coord\_flip() +

scale\_color\_manual(values=c("#FB8072","#80B1D3"))+

scale\_fill\_manual(values=c("#FB8072","#80B1D3"))+

geom\_signif(comparisons = my\_comparisons,

step\_increase = 0.5,

map\_signif\_level = F,

test = wilcox.test,size=0.5,textsize = 3)+

theme\_classic()

pdf(file="IPS.pdf", width=4.5, height=2)

p=p+rotate\_y\_text(60)

print(p)

dev.off()

22. ####### Boxplot of neoantigens (Figure 7B) #######

Upload the file “NeoAgs.txt” to online website ([www.xiantao.love](http://www.xiantao.love)) and download the violin plot (Figure 7B).

23. ####### Boxplot of TIDE score (Figure 7C, D and E) #######

Upload the file “TIDE\_riskscore.txt” to online website ([www.xiantao.love](http://www.xiantao.love)) and download the scatterplot (Figure 7C).

Upload the file “TIDE\_risk.txt” to online website ([www.xiantao.love](http://www.xiantao.love)) and download the dotplot (Figure 7D).

Upload the file “TIDE.responde\_risk.txt” to online website ([www.xiantao.love](http://www.xiantao.love)) and download the stacked barplot (Figure 7E).

###################Validation in IMvigor210 cohort #######################

24.####### Survival curse (Figure 7F) #######

library(survival)

library(caret)

library(glmnet)

library(survminer)

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

multiCox <- coxph(Surv(futime, fustat) ~ ., data = rt)

multiCox=step(multiCox,direction = "both")

multiCoxSum=summary(multiCox)

score=predict(multiCox, type="risk", newdata=rt)

coxGene=rownames(multiCoxSum$coefficients)

coxGene=gsub("`", "", coxGene)

vigor=read.table("IMvigor210\_exp.txt", header=T, sep="\t", check.names=F, row.names=1)

vigor=t(vigor[coxGene,])

cli=read.table("IMvigor210\_time.txt", header=T, sep="\t", check.names=F, row.names=1)

colnames(cli)=c("futime", "fustat")

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

vigorTime=cbind(cli[sameSample,,drop=F], vigor[sameSample,,drop=F])

#vigorTime[,3:ncol(vigorTime)]=vigorTime[,3:ncol(vigorTime)]\*median(as.matrix(rt[,coxGene]))/median(as.matrix(vigorTime[,3:ncol(vigorTime)]))

vigorScore=predict(multiCox, type="risk", newdata=vigorTime)

Risk=as.vector(ifelse(vigorScore>median(vigorScore), "high", "low"))

vigorRiskOut=cbind(vigorTime, riskScore=as.vector(vigorScore), Risk)

vigorRiskOut=cbind(id=rownames(vigorRiskOut), vigorRiskOut)

write.table(vigorRiskOut,file="risk.IMvigor.txt",sep="\t",quote=F,row.names=F)

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

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

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=paste("Log-rank test",pValue),

pval.size=5,

title="IMvigor210",

legend = c(0.75, 0.8),

legend.title="Risk",

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

xlab="Time(years)",

font.title = c(15, "bold", "darkblue"),

font.x=c(13, "bold.italic", "#FF4040"),

font.y=c(13, "bold.italic", "darkred"),

font.legend=12,

font.tickslab=12,

risk.table.fontsize = 4,

risk.table.font.x=12,

risk.table.font.tickslab=10,

break.time.by = 2,

palette=c("#FB8072","#80B1D3"),

risk.table=TRUE,

tables.y.text = FALSE,

cumevents=F,

ggtheme = theme\_classic()+

theme(plot.title = element\_text(size=12,hjust=0.5)),

risk.table.col = "strata",

risk.table.height=.25)

pdf(file=outFile,onefile = FALSE,width = 5,height =5)

print(surPlot)

dev.off()

}

bioSurvival(inputFile="risk.IMvigor.txt", outFile="sur.IMvigor.pdf")

25. ####### Boxplot of TIDE score (Figure 7G and H) #######

Upload the file “response\_riskscore.txt” to online website ([www.xiantao.love](http://www.xiantao.love)) and download the dotplot (Figure 7G).

Upload the file “responde\_risk.txt” to online website ([www.xiantao.love](http://www.xiantao.love)) and download the stacked barplot (Figure 7H).