Supplementary File 1. The code script of the bioinformatical analysis.

#01-download data from TCGA

library(TCGAbiolinks)

library(tidyverse)

library(SummarizedExperiment)

setwd("01.data")

#mRNA profile

count\_lusc <- GDCquery(project = "TCGA-LUSC",

data.category = "Transcriptome Profiling",

data.type = "Gene Expression Quantification",

workflow.type = "HTSeq - Counts")

)

GDCdownload(count\_lusc)

exp\_count\_lusc <- GDCprepare(query = count\_lusc,save = TRUE, save.filename = "count\_lusc.rda")

fpkm\_lusc <- GDCquery(project = "TCGA-LUSC",

data.category = "Transcriptome Profiling",

data.type = "Gene Expression Quantification",

workflow.type = "HTSeq - FPKM-UQ")

GDCdownload(fpkm\_lusc)

exp\_fpkm\_lusc <- GDCprepare(query = fpkm\_lusc,save = TRUE, save.filename = "fpkm\_lusc.rda")

count\_luad <- GDCquery(project = "TCGA-LUAD",

data.category = "Transcriptome Profiling",

data.type = "Gene Expression Quantification",

workflow.type = "HTSeq - Counts")

)

GDCdownload(count\_luad)

exp\_count\_luad <- GDCprepare(query = count\_luad,save = TRUE, save.filename = "count\_luad.rda")

fpkm\_luad <- GDCquery(project = "TCGA-LUAD",

data.category = "Transcriptome Profiling",

data.type = "Gene Expression Quantification",

workflow.type = "HTSeq - FPKM-UQ")

GDCdownload(fpkm\_luad)

exp\_fpkm\_luad <- GDCprepare(query = fpkm\_luad,save = TRUE, save.filename = "fpkm\_luad.rda")

#miRNA

mirna\_lusc <- GDCquery(project = "TCGA-LUSC",

experimental.strategy = "miRNA-Seq",

data.category = "Transcriptome Profiling",

# barcode = c("TARGET-20-PATDNN","TARGET-20-PAPUNR"),

data.type = "miRNA Expression Quantification")

GDCdownload(mirna\_lusc)

mirna\_lusc <- GDCprepare(query = mirna\_lusc,save = TRUE, save.filename = "mirna\_lusc.rda")

mirna\_luad <- GDCquery(project = "TCGA-LUAD",

experimental.strategy = "miRNA-Seq",

data.category = "Transcriptome Profiling",

# barcode = c("TARGET-20-PATDNN","TARGET-20-PAPUNR"),

data.type = "miRNA Expression Quantification")

GDCdownload(mirna\_luad)

mirna\_luad <- GDCprepare(query = mirna\_luad,save = TRUE, save.filename = "mirna\_luad.rda")

#mutation profile

maf\_lusc <- GDCquery\_Maf("LUSC", pipelines = "muse")

saveRDS(maf\_lusc,file="mal\_lusc.rds")

maf\_lusc <-readRDS("mal\_lusc.rds")

maf\_luad <- GDCquery\_Maf("LUAD", pipelines = "muse")

saveRDS(maf\_luad,file="mal\_luad.rds")

maf\_luad <-readRDS("mal\_luad.rds")

#clinic information

clin\_lusc <- GDCquery\_clinic("TCGA-LUSC", "clinical")

write.table(clin\_lusc,"clin\_lusc.tsv",sep = '\t',row.names = F)

clin\_lusc <- read.table("clin\_lusc.tsv",sep = '\t',header = T)

clin\_luad <- GDCquery\_clinic("TCGA-LUAD", "clinical")

write.table(clin\_luad,"clin\_luad.tsv",sep = '\t',row.names = F)

clin\_luad <- read.table("clin\_luad.tsv",sep = '\t',header = T)

#get sample type

samplesDown <- getResults(count\_lusc,cols=c("cases"))

saveRDS(samplesDown,file = "lusc\_sampledown.rds")

dataSmTP <- TCGAquery\_SampleTypes(barcode = samplesDown,

typesample = "TP")

dataSmNT <- TCGAquery\_SampleTypes(barcode = samplesDown,

typesample = "NT")

#02-hc-cluster

library(tidyverse)

library(reshape2)

#mutation file: maf\_lusc, maf\_luad

#clinical file: clin\_lusc, clin\_luad

#public markers

markers\_pub<- c("EGFR","ALK","KARS","TP53","PIK3CA","BRAF","MET","NF1")

markers\_pub <- c("EGFR","BRAF","ROS1","FGFR1","KRAS","MET","HER2","NRTK","RET","ALK",

"NRG1","TP53","PTEN","PIK3CA")

library(openxlsx)

setwd("analysis/01.data")

markers3 <- read.xlsx("NIHMS948705-supplement-8.xlsx",sheet = 2,colNames = T)

markers3%>%

filter(X2=="LUSC")->marker.lusc

marker.lusc$`Table.S1:.Final.gene.consensus.list..Related.to.Figure.2.`->marker.lusc.1

markers3%>%

filter(X2=="LUAD")->marker.luad

marker.luad$`Table.S1:.Final.gene.consensus.list..Related.to.Figure.2.`->marker.luad.1

marker.lung <- unique(c(marker.lusc.1,marker.luad.1))

#path

setwd("analysis/02.cluster\_by\_mutation")

unique(maf\_lusc$Tumor\_Sample\_Barcode)[!(grepl("-01A-",unique(maf\_lusc$Tumor\_Sample\_Barcode)))]

length(unique(maf\_lusc$Hugo\_Symbol))

#1

quantile(table(maf\_lusc$Hugo\_Symbol))

#2

maf\_lusc%>%

dplyr::select(Hugo\_Symbol,Chromosome,Start\_Position,End\_Position,Strand,

Variant\_Classification,

Tumor\_Sample\_Barcode)%>%

filter(Hugo\_Symbol %in% marker.lusc.1)%>%

mutate(var\_id = paste0(Hugo\_Symbol,"\_",

Chromosome,"\_",

Start\_Position,"\_",

End\_Position,"\_",

Strand))->maf\_lusc\_brief

dim(maf\_lusc\_brief)

maf\_luad%>%

dplyr::select(Hugo\_Symbol,Chromosome,Start\_Position,End\_Position,Strand,

Variant\_Classification,

Tumor\_Sample\_Barcode)%>%

filter(Hugo\_Symbol %in% marker.luad.1)%>%

mutate(var\_id = paste0(Hugo\_Symbol,"\_",

Chromosome,"\_",

Start\_Position,"\_",

End\_Position,"\_",

Strand))->maf\_luad\_brief

dim(maf\_luad\_brief)

maf\_lusc%>%

rbind(maf\_luad)%>%

select(Hugo\_Symbol,Chromosome,Start\_Position,End\_Position,Strand,

Variant\_Classification,

Tumor\_Sample\_Barcode)%>%

filter(Hugo\_Symbol %in% marker.lung)%>%

mutate(var\_id = paste0(Hugo\_Symbol,"\_",

Chromosome,"\_",

Start\_Position,"\_",

End\_Position,"\_",

Strand))->maf\_lung\_brief

dim(maf\_lung\_brief)

#table(maf\_lusc\_brief$Hugo\_Symbol)[order(table(maf\_lusc\_brief$Hugo\_Symbol),decreasing = T)][1:50]->features

#write.table(as.data.frame(features),file = "selected\_mutation\_genes.csv",sep = ',')

#features <- read.table("selected\_mutation\_genes.csv",sep = ',')

#table(maf\_lusc\_brief$var\_id)

maf\_lusc\_brief%>%

dplyr::select(Tumor\_Sample\_Barcode,Hugo\_Symbol)%>%

mutate(count=1)->test

dcast(test,Hugo\_Symbol~Tumor\_Sample\_Barcode)->maf\_lusc\_df

rownames(maf\_lusc\_df) <-maf\_lusc\_df$Tumor\_Sample\_Barcode

maf\_luad\_brief%>%

dplyr::select(Tumor\_Sample\_Barcode,Hugo\_Symbol)%>%

mutate(count=1)->test

dcast(test,Hugo\_Symbol~Tumor\_Sample\_Barcode)->maf\_luad\_df

rownames(maf\_luad\_df) <-maf\_luad\_df$Tumor\_Sample\_Barcode

maf\_lung\_brief%>%

select(Tumor\_Sample\_Barcode,Hugo\_Symbol)%>%

mutate(count=1)->test

dcast(test,Hugo\_Symbol~Tumor\_Sample\_Barcode)->maf\_lung\_df

rownames(maf\_lung\_df) <-maf\_lung\_df$Tumor\_Sample\_Barcode

maf\_lusc\_df[is.na(maf\_lusc\_df)] <-0

maf\_lusc\_df\_tmp <- apply(maf\_lusc\_df[,-1], 2,as.numeric)

rownames(maf\_lusc\_df\_tmp) <-maf\_lusc\_df$Hugo\_Symbol

maf\_luad\_df\_tmp <- apply(maf\_luad\_df[,-1], 2,as.numeric)

rownames(maf\_luad\_df\_tmp) <-maf\_luad\_df$Hugo\_Symbol

maf\_lung\_df\_tmp <- apply(maf\_lung\_df[,-1], 2,as.numeric)

rownames(maf\_lung\_df\_tmp) <-maf\_lung\_df$Hugo\_Symbol

saveRDS(maf\_lusc\_df\_tmp,"maf\_lusc.df.rds")

maf\_lusc\_df\_tmp <- readRDS("maf\_lusc.df.rds")

dim(maf\_lusc\_df\_tmp)

saveRDS(maf\_luad\_df\_tmp,"maf\_luad.df.rds")

maf\_lusc\_df\_tmp <- readRDS("maf\_luad.df.rds")

head(maf\_luad\_df\_tmp)[,1:6]

#

library(pheatmap)

maf\_lusc\_df\_tmp->hp\_df

pdf(file = "lusc/hp.drvergene.mutfreq.pdf",width = 8,height = 6)

pheatmap(hp\_df)

dev.off()

maf\_luad\_df\_tmp->hp\_df

pdf(file = "luad/hp.drvergene.mutfreq.pdf",width = 8,height = 6)

pheatmap(hp\_df)

dev.off()

#

#maf <- scale(t(maf\_lusc\_df\_tmp))

#

maf <- scale(t(maf\_luad\_df\_tmp))

hc2<-hclust(dist(maf),method = "single")

pdf("hc.single.pdf",width = 8,height = 12)

plot(hc2,hang = -0.01,cex=0.7)

dev.off()

library(NbClust)

nc<-NbClust(maf,distance = "euclidean",

min.nc = 2,max.nc = 20,

method = "ward.D2")

table(nc$Best.nc[1,])

barplot(table(nc$Best.nc[1,]))

clusters<-cutree(hc,k=18)

table(clusters)

write.table(as.data.frame(clusters),file = "luad/m3.clusters.csv",sep = ',')

pdf("hc.m3.cluster.pdf",width = 8,height = 4)

plot(hc,cex=0.7,hang=-0.01)

rect.hclust(hc,k=5,border="red")

dev.off()

clusters <- read.table("luad/m3.clusters.csv",sep = ',')

clusters$barcode <- rownames(clusters)

table(clusters$clusters)

#for combine

clusters$clusters <- as.character(clusters$clusters)

clusters%>%

mutate(cancer\_type = ifelse(barcode %in% maf\_lusc$Tumor\_Sample\_Barcode,"LUSC","LUAD"))%>%

select(-barcode)->clusters

clusters[order(clusters$clusters),]->clusters

clusters%>%

dplyr::select(-barcode)->clusters

pdf("luad/hp.gcluster.pdf",width = 8,height = 6)

pheatmap(maf\_luad\_df\_tmp[,rownames(clusters)],annotation\_col = clusters,cluster\_cols = F)

dev.off()

pdf("luad/hp.acluster.pdf",width = 8,height = 7)

pheatmap(maf\_luad\_df\_tmp[,rownames(clusters)],annotation\_col = clusters,cluster\_cols = T)

dev.off()

#03-cluster-survival

setwd("analysis/02.cluster\_by\_mutation")

clusters <- read.table(file = "lusc/m3.clusters.csv",sep = ',')

clusters$clusters<- as.character(clusters$clusters)

clusters$sample\_id <- rownames(clusters)

clusters[order(clusters$clusters),]->clusters

clusters%>%

select(-sample\_id)->clusters

clusters["submitter\_id"] <- str\_sub(rownames(clusters),1,12)

setwd("/analysis/03.cluster\_clincal")

maf\_lusc%>%

select(Hugo\_Symbol,Tumor\_Sample\_Barcode)%>%

mutate(count=1)->maf\_lusc\_brief

maf\_lusc\_brief%>%

mutate(clusters=clusters[maf\_lusc\_brief$Tumor\_Sample\_Barcode,])->maf\_cluster

table(maf\_cluster$clusters)

#mutation

#as.data.frame(table(maf\_cluster$Hugo\_Symbol,maf\_cluster$clusters))-> mut\_clu

#dcast(mut\_clu,Var1~Var2,value.var = "Freq")->mut\_clu

#rownames(mut\_clu)<-mut\_clu$Var1

view(head(maf\_lusc))

dcast(maf\_lusc\_brief,Hugo\_Symbol~Tumor\_Sample\_Barcode) -> mut\_df

rownames(mut\_df) <- mut\_df$Hugo\_Symbol

head(mut\_df)[,1:6]

dim(mut\_df)

library(pheatmap)

mut\_df[rowSums(mut\_df[,-1])>20,]->mut\_df\_filt

dim(mut\_df\_filt)

mut\_df -> mut\_df\_filt

pdf("hp.m3.clu.pdf",width = 6,height = 6)

pheatmap(mut\_df\_filt[marker.lusc.1,rownames(clusters)],

annotation\_col = clusters,

cluster\_cols=F)

dev.off()

pheatmap()

#clinical

clin\_lusc

colnames(clin\_lusc)

clusters$submitter\_id <- substr(rownames(clusters),1,12)

clin\_lusc%>%

select(submitter\_id,ajcc\_pathologic\_stage)%>%

right\_join(clusters)->clin\_stage

dim(clin\_stage)

head(clin\_stage)

clin\_luad%>%

select(submitter\_id,ajcc\_pathologic\_stage)%>%

right\_join(clusters)->clin\_stage

ggplot(clin\_stage,aes(clusters,fill=ajcc\_pathologic\_stage))+

geom\_bar(stat="count")+

theme\_bw()->p

ggsave(filename = "m3.stage.cluster.pdf",plot = p, width = 6,height = 6)

#survival

data.frame(submitter\_id = substr(colnames(maf\_lusc\_df\_tmp),1,12),

clusters = ifelse(maf\_lusc\_df\_tmp[i,]>0,1,0))->clusters

setwd("analysis/01.data")

clin\_lusc <- read.table("clin\_lusc.tsv",sep = '\t',header = T)

clin\_lusc%>%

select(submitter\_id,days\_to\_last\_follow\_up,days\_to\_death,vital\_status)%>%

left\_join(clusters)->sur\_df

clin\_luad%>%

select(submitter\_id,days\_to\_last\_follow\_up,days\_to\_death,vital\_status)%>%

right\_join(clusters)->sur\_df

sur\_df$days\_to\_death[is.na(sur\_df$days\_to\_death)] <- 0

sur\_df$days\_to\_last\_follow\_up[is.na(sur\_df$days\_to\_last\_follow\_up)] <- 0

sur\_df$days=as.numeric(sur\_df$days\_to\_death)+as.numeric(sur\_df$days\_to\_last\_follow\_up)

#sur\_df$days=round(sur\_df$days/30,2)

1library(survminer)

library(survival)

plot\_2g\_sur <- function(g1,g2,sur\_df){

sur\_df%>%

filter(clusters %in% c(g1,g2))->sur\_df\_clu

TCGAanalyze\_survival(sur\_df\_clu,

clusterCol="clusters",

legend = "clusters",

risk.table = FALSE,

conf.int = FALSE,

color = c("Dark2"),

#main = gene\_symbol,

xlab = "Time since diagnosis (Days)",

dpi=1200,

height = 4, width = 5,

#dpi = 300,

filename = paste0("luad/m3.clu.",g1,".",g2,".survival.pdf"))

}

plot\_2g\_sur("1","2",sur\_df)

...

plot\_2g\_sur("17","18",sur\_df)

#04-TILsig-ceRNA

#TILSig

#immune-related mRNA

setwd("analysis/01.data/GEO-immune-subset")

exp<- read.table("GSE28490\_series\_matrix.d.txt",header = T)

soft <- readr::read\_tsv("GSE28490\_family.soft.d1.txt",col\_names = T)

soft%>%

select(ID,`Gene Symbol`)->soft\_b

cell\_type <- read.table("GSE28490.cell\_type")

colnames(cell\_type) <- colnames(exp)

data.frame(t(cell\_type)[-1,])->celltype

for(ct in unique(celltype$t.cell\_type...1...)){

ct = unique(celltype$t.cell\_type...1...)[9]

sample\_list <- rownames(celltype)[celltype$t.cell\_type...1...==ct]

exp[,c("ID\_REF",sample\_list)]%>%

left\_join(soft\_b,by = c("ID\_REF" = "ID"))%>%

filter(!is.na(`Gene Symbol`))%>%

select(-ID\_REF)->exp\_tmp

exp\_tmp$exp\_mean <- apply(exp\_tmp[,1:length(exp\_tmp)-1],1,mean)

exp\_tmp[order(exp\_tmp$exp\_mean,decreasing = T),]->sig\_mRNA

sig\_mRNA[1:floor(nrow(exp\_tmp)\*0.05),]->df\_ct9

}

intersect(df\_ct1$`Gene Symbol`,c("PTPRC"))

Reduce(intersect,list(v1 = df\_ct1$`Gene Symbol`,

v2 = df\_ct2$`Gene Symbol`,

v3 = df\_ct3$`Gene Symbol`,

v4 = df\_ct4$`Gene Symbol`,

v5 = df\_ct5$`Gene Symbol`,

v6 = df\_ct6$`Gene Symbol`,

v7 = df\_ct7$`Gene Symbol`,

v8 = df\_ct8$`Gene Symbol`,

v9 = df\_ct9$`Gene Symbol`))->immun.mRNA

setwd("analysis/04.TILsig-ceRNA")

write.table(as.data.frame(immun.mRNA),file = "mRNA.signature",sep=',')

#immune-related miRNA

exp\_mi<- read.table("GSE28487\_series\_matrix.d.txt",header = T)

soft\_mi <- readr::read\_tsv("GSE28487\_family.soft.txt",col\_names = F)

soft\_mi%>%

select(X1,X8)->soft\_mi\_b

cell\_type\_mi <- read.table("GSE28487.cell\_type")

colnames(cell\_type\_mi) <- colnames(exp\_mi)

data.frame(t(cell\_type\_mi)[-1,])->celltype\_mi

for(ct in unique(celltype\_mi$t.cell\_type\_mi...1...)){

ct = unique(celltype\_mi$t.cell\_type\_mi...1...)[1]

sample\_list <- rownames(celltype\_mi)[celltype\_mi$t.cell\_type\_mi...1...==ct]

exp\_mi[,c("ID\_REF",sample\_list)]%>%

left\_join(soft\_mi\_b,by = c("ID\_REF" = "X1"))%>%

filter(!is.na(X8))%>%

select(-ID\_REF)->exp\_mi\_tmp

exp\_mi\_tmp$exp\_mean <- apply(exp\_mi\_tmp[,1:length(exp\_mi\_tmp)-1],1,mean)

exp\_mi\_tmp[order(exp\_mi\_tmp$exp\_mean,decreasing = T),]->sig\_miRNA

sig\_miRNA[1:floor(nrow(exp\_mi\_tmp)\*0.05),]->dfmi\_ct1

}

Reduce(intersect,list(v1 = dfmi\_ct1$X8,

v2 = dfmi\_ct2$X8,

v3 = dfmi\_ct3$X8,

v4 = dfmi\_ct4$X8,

v5 = dfmi\_ct5$X8,

v6 = dfmi\_ct6$X8,

v7 = dfmi\_ct7$X8,

v8 = dfmi\_ct8$X8,

v9 = dfmi\_ct9$X8))->immun.miRNA

setwd("C:/Users/FuXin/OneDrive/FU/LiKe/LungCancer-202112-202206/analysis/04.TILsig-ceRNA")

write.table(as.data.frame(immun.miRNA),file = "miRNA.signature",sep=',')

#lncRNA

library(openxlsx)

immun.linc <- read.xlsx("jitc-2019-000110supp001-lnc-pubmed.xlsx",sheet = 1,startRow = 1,colNames = T)

write.table(as.data.frame(immun.linc),file = "lncRNA.signature",sep=',')

library(pheatmap)

head(exp)[,1:6]

head(soft\_b)

soft\_b%>%

filter(`Gene Symbol` %in% immun.mRNA)%>%

left\_join(exp,by = c("ID" = "ID\_REF"))%>%

select(-ID)->hp\_df

hp\_df[is.na(hp\_df)]<-0

hp\_df[is.infinite(hp\_df)]<-0

colnames(celltype) <- "Immune cell type"

for (i in 1:nrow(celltype)) {

#print(i)

celltype$t.cell\_type...1...[i] <- unlist(strsplit(celltype$t.cell\_type...1...,split = ':')[[i]][2])

}

hp\_df[rowSums(hp\_df[,-1])>0,]->hp\_df

hp\_df\_uniq <- data.frame()

for (gene in unique(hp\_df$`Gene Symbol`)) {

hp\_df%>%

filter(`Gene Symbol`==gene)->tmp

hp\_df\_uniq%>%

rbind(tmp[1,])->hp\_df\_uniq

}

as.data.frame(hp\_df\_uniq)->hp\_df\_uniq

rownames(hp\_df\_uniq) <- hp\_df\_uniq$`Gene Symbol`

pdf("hp.immune.mRNA.pdf",width = 14,height = 16)

pheatmap(hp\_df\_uniq[,-1],

annotation\_col = celltype,

)

dev.off()

#miRNA

soft\_mi\_b%>%

filter(X8 %in% immun.miRNA)%>%

left\_join(exp\_mi,by = c("X1" = "ID\_REF"))%>%

select(-X1)->hp\_df\_mi

hp\_df\_mi[is.na(hp\_df\_mi)]<-0

colnames(celltype\_mi) <- "Immune cell type"

for (i in 1:nrow(celltype\_mi)) {

#print(i)

celltype\_mi$t.cell\_type\_mi...1...[i] <- unlist(strsplit(celltype\_mi$t.cell\_type\_mi...1...,split = ':')[[i]][2])

}

hp\_df\_mi[rowSums(hp\_df\_mi[,-1])>0,]->hp\_df\_mi

hp\_df\_mi\_uniq <- data.frame()

for (gene in unique(hp\_df\_mi$X8)) {

hp\_df\_mi%>%

filter(X8==gene)->tmp

hp\_df\_mi\_uniq%>%

rbind(tmp[1,])->hp\_df\_mi\_uniq

}

as.data.frame(hp\_df\_mi\_uniq)->hp\_df\_mi\_uniq

rownames(hp\_df\_mi\_uniq) <- hp\_df\_mi\_uniq$X8

pdf("hp.immune.miRNA.pdf",width = 10,height = 5)

pheatmap(hp\_df\_mi\_uniq[,-1],

annotation\_col = celltype\_mi,

)

dev.off()

#05-SigRNA-exp-clusters

library(tidyverse)

library(pheatmap)

library(ggsignif)

setwd("analysis/04.TILsig-ceRNA")

#signature:immun.mRNA, immun.miRNA, immun.linc

immun.mRNA<-read.table("mRNA.signature",sep = ',')$immun.mRNA

immun.miRNA<-read.table("miRNA.signature",sep = ',')$immun.miRNA

immun.linc<-read.table("lncRNA.signature",sep = ',')$immun.linc

sig\_mirna <- list()

for (mi in immun.miRNA) {

sig\_mirna <-c(unlist(strsplit(mi,split = '//')),sig\_mirna)

}

sig\_mirna <- unlist(sig\_mirna)

sig\_mrna <- list()

for (mg in immun.mRNA) {

sig\_mrna <-c(unlist(strsplit(mg,split = '//')[[1]][1]),sig\_mrna)

}

sig\_mrna <- unlist(sig\_mrna)

sig\_lncrna <- immun.linc$lncRNAs

#cluster

setwd("analysis/02.cluster\_by\_mutation")

clusters <- read.table("luad/m3.clusters.csv",sep = ',')

clusters$submitter\_id <- substr(rownames(clusters),1,12)

clusters$clusters<- as.character(clusters$clusters)

clusters[order(clusters$clusters),]->clusters

#data

setwd("analysis/01.data")

load("mirna\_lusc.rda")

miRNA\_exp <- data

mRNA\_exp <- load("fpkm\_lusc.rda")

mRNA\_exp <- data

load("mirna\_luad.rda")

miRNA\_exp <- data

mRNA\_exp <- load("fpkm\_luad.rda")

mRNA\_exp <- data

#miRNA expr

setwd("analysis/05.SigceRNA-exp-cluster")

mirna\_exp\_clu <- data.frame(miRNA\_ID = miRNA\_exp$miRNA\_ID)

for (i in 1:nrow(clusters)) {

cll <- grepl(paste0("reads\_per\_million\_miRNA\_mapped\_",clusters[i,]$submitter\_id),colnames(miRNA\_exp))

df\_tmp <-data.frame(C1 = miRNA\_exp[,colnames(miRNA\_exp)[cll]])

colnames(df\_tmp) <-colnames(miRNA\_exp)[cll]

mirna\_exp\_clu%>%

cbind(df\_tmp)->mirna\_exp\_clu

}

rownames(miRNA\_exp) <- miRNA\_exp$miRNA\_ID

intersect(miRNA\_exp$miRNA\_ID,sig\_mirna)->sig\_mirna\_1

miRNA\_exp[sig\_mirna\_1,colnames(mirna\_exp\_clu)] -> sig\_mirna\_exp

data.frame(sample=colnames(mirna\_exp\_clu),

submitter\_id = substr(colnames(mirna\_exp\_clu),32,43))%>%

left\_join(clusters)->colData\_mirna

rownames(colData\_mirna)<- colData\_mirna$sample

colData\_mirna%>%

dplyr::select(clusters)->colData\_mirna

colData\_mirna%>%

filter(!is.na(clusters)) -> colData\_mirna

library(pheatmap)

pdf("hp.luad.miRNA.signature.TCGA.pdf",width = 8,height = 8)

pheatmap(sig\_mirna\_exp[,-1],

annotation\_col = colData\_mirna,

cluster\_cols = F)

dev.off()

for (i in 1:nrow(sig\_mirna\_exp)) {

data.frame(t(sig\_mirna\_exp[i,])[-1,])->exp\_tmp

exp\_tmp$sample <- rownames(exp\_tmp)

exp\_tmp%>%

left\_join(colData\_mirna)->exp\_tmp1

exp\_tmp1$expression <- as.numeric(exp\_tmp1$t.sig\_mirna\_exp.i......1...)

ggplot(exp\_tmp1%>%

dplyr::select(-sample),aes(x=clusters,y=expression))+

ylab(label = rownames(sig\_mirna\_exp)[i])+

geom\_boxplot(fill = color\_cus[i+150])+

geom\_signif(comparisons = com\_luad,

map\_signif\_level=T,

test=wilcox.test,step\_increase=0.2)+

theme\_bw()->p

ggsave(filename = paste0("luad\_miRNA/",rownames(sig\_mirna\_exp)[i],".pdf"),

width = 7,height = 3)

}

colData\_mirna$sample <- rownames(colData\_mirna)

color\_cus <- colors()

dev.off()

#mRNA

#immunecellAI

clusters%>%

filter(clusters=="1")->cdf

sample\_list = list()

for(sample in rownames(clusters)){

bc = substr(sample,1,16)

sample\_add <- colnames(mRNA\_exp)[grepl(bc,colnames(mRNA\_exp))]

sample\_list <- unlist(c(sample\_list,sample\_add))

}

mRNA\_exp[,sample\_list]->cluster1\_exp

c1\_exp\_fpkm <- cluster1\_exp@assays@data$`HTSeq - FPKM-UQ`

rownames(c1\_exp\_fpkm) <-rownames(cluster1\_exp)

colnames(c1\_exp\_fpkm) <- colnames(cluster1\_exp)

write.table(c1\_exp\_fpkm,

file = "cluster1.lusc.exp.txt",

sep = '\t')

genelist <- mRNA\_exp@rowRanges$ensembl\_gene\_id[mRNA\_exp@rowRanges$external\_gene\_name %in% sig\_mrna]

mRNA\_exp[genelist,sample\_list] ->sig\_mrna\_exp

sig\_mrna\_exp\_fpkm <- sig\_mrna\_exp@assays@data$`HTSeq - FPKM-UQ`

rownames(sig\_mrna\_exp\_fpkm) <-rownames(sig\_mrna\_exp)

colnames(sig\_mrna\_exp\_fpkm) <- colnames(sig\_mrna\_exp)

data.frame(sample = sample\_list,submitter\_id = substr(sample\_list,1,12))%>%

left\_join(clusters)%>%

dplyr::select(clusters)->colData\_mrna

fpkmToTpm <- function(fpkm)

{

exp(log(fpkm) - log(sum(fpkm)) + log(1e6))

}

sig\_mrna\_exp\_tpm <- as.data.frame (apply(sig\_mrna\_exp\_fpkm, 2, fpkmToTpm))

rownames(colData\_mrna) <- sample\_list

pdf(file = "hp.luad.mRNA.signature.TCGA.pdf",width = 8,height = 8)

pheatmap(sig\_mrna\_exp\_tpm,annotation\_col = colData\_mrna,cluster\_cols = F)

dev.off()

colData\_mrna$sample <- rownames(colData\_mrna)

com\_lusc <- list(c("1", "3"),

c("2","13"),

c("3","16"),

c("9","13"),

c("15","13"))

com\_luad <- list(c("1","9"),c("5","12"),c("6","7"),c("7","9"),c("8","12"),

c("9","12"),c("12","14"),c("12","17"))

for (i in 1:nrow(sig\_mrna\_exp\_tpm)) {

genename <- mRNA\_exp@rowRanges$external\_gene\_name[mRNA\_exp@rowRanges$ensembl\_gene\_id==rownames(sig\_mrna\_exp\_tpm)[i]]

data.frame(t(sig\_mrna\_exp\_tpm[i,])[-1,])->exp\_tmp

exp\_tmp$sample <- rownames(exp\_tmp)

exp\_tmp%>%

left\_join(colData\_mrna)->exp\_tmp1

exp\_tmp1$expression <- as.numeric(exp\_tmp1$t.sig\_mrna\_exp\_tpm.i......1...)

ggplot(exp\_tmp1%>%

dplyr::select(-sample),aes(x=clusters,y=expression))+

ylab(label = genename)+

geom\_boxplot(fill = color\_cus[i+105])+

geom\_signif(comparisons = com\_luad,

map\_signif\_level=T,

test=wilcox.test,step\_increase=0.2)+

theme\_bw()->p

ggsave(filename = paste0("luad\_mRNA/",genename,".pdf"),

width = 7,height = 3)

}

#lncRNA

lncrna\_genelist1 <- mRNA\_exp@rowRanges$ensembl\_gene\_id[mRNA\_exp@rowRanges$external\_gene\_name %in% sig\_lncrna]

lncrna\_genelist2 <- intersect(mRNA\_exp@rowRanges$ensembl\_gene\_id,sig\_lncrna)

lncrna\_exp <- mRNA\_exp[c(lncrna\_genelist1,lncrna\_genelist2),sample\_list]

lncrna\_exp\_fpkm <- lncrna\_exp@assays@data$`HTSeq - FPKM-UQ`

rownames(lncrna\_exp\_fpkm) <- rownames(lncrna\_exp)

colnames(lncrna\_exp\_fpkm) <- colnames(lncrna\_exp)

data.frame(sample = sample\_list,submitter\_id = substr(sample\_list,1,12))%>%

left\_join(clusters)%>%

dplyr::select(clusters)->colData\_lncrna

rownames(colData\_lncrna) <- sample\_list

lncrna\_exp\_tpm <- as.data.frame (apply(lncrna\_exp\_fpkm, 2, fpkmToTpm))

dim(lncrna\_exp\_tpm)

pdf(file = "hp.luad.lncRNA.signature.col.TCGA.pdf",width = 8,height = 8)

pheatmap(lncrna\_exp\_tpm,annotation\_col = colData\_lncrna,cluster\_cols = T)

dev.off()

colData\_lncrna$sample <- rownames(colData\_lncrna)

for (i in 1:nrow(lncrna\_exp\_tpm)) {

genename <- mRNA\_exp@rowRanges$external\_gene\_name[mRNA\_exp@rowRanges$ensembl\_gene\_id==rownames(sig\_mrna\_exp\_tpm)[i]]

data.frame(t(lncrna\_exp\_tpm[i,])[-1,])->exp\_tmp

exp\_tmp$sample <- rownames(exp\_tmp)

exp\_tmp%>%

left\_join(colData\_mrna)->exp\_tmp1

exp\_tmp1$expression <- as.numeric(exp\_tmp1$t.lncrna\_exp\_tpm.i......1...)

ggplot(exp\_tmp1%>%

dplyr::select(-sample),aes(x=clusters,y=expression))+

ylab(label = genename)+

geom\_boxplot(fill = color\_cus[i+105])+

geom\_signif(comparisons = com\_luad,

map\_signif\_level=T,

test=wilcox.test,step\_increase=0.2)+

theme\_bw()->p

ggsave(filename = paste0("luad\_lncRNA/",genename,".pdf"),

width = 7,height = 3)

}

#06-Immune-features

library(reshape2)

library(ggplot2)

setwd("analysis/06.cluster-immunescore")

#immunecellAI

lusc\_immai <- readr::read\_tsv("immunecellai/LUSC\_self.txt",col\_names = T,)

lusc\_immai$sample <- gsub("\\.","-",lusc\_immai$sample)

luad\_immai <- readr::read\_tsv("immunecellai/LUAD\_self.txt",col\_names = T)

luad\_immai$sample <- gsub("\\.","-",luad\_immai$sample)

setwd("analysis/02.cluster\_by\_mutation")

clusters <- read.table("luad/m3.clusters.csv",sep = ',')

clusters$ID <- substr(rownames(clusters),1,15)

clusters$clusters <- as.character(clusters$clusters)

clusters$sample\_s <- clusters$ID

lusc\_immai$sample\_s <- substr(lusc\_immai$sample,1,15)

luad\_immai$sample\_s <- substr(luad\_immai$sample,1,15)

clusters%>%

left\_join(luad\_immai,by = c("sample\_s" = "sample\_s"))%>%

dplyr::select(-sample,-sample\_s,-ID)->immai\_df

rownames(immai\_df) <- rownames(clusters)

dim(immai\_df)

dim(clusters)

for(i in 2:length(immai\_df)){

cell\_type = colnames(immai\_df)[i]

df<-immai\_df[,c(1,i)]

colnames(df) <- c("clusters","celltype")

ggplot(df,aes(x=clusters,y=celltype))+ylab(label = cell\_type)+

geom\_boxplot(fill = color\_cus[i+10])+theme\_bw()->p

ggsave(filename = paste0("immunecellai/dist.luad.",cell\_type,".pdf"),p,

width = 8,height = 1.2)

}

color\_cus <- colors()

#estimate score

lusc\_ssis <- readr::read\_tsv("estimate\_score/lung\_squamous\_cell\_carcinoma\_RNAseqV2.txt",col\_names = T)

luad\_ssis <- readr::read\_tsv("estimate\_score/lung\_adenocarcinoma\_RNAseqV2.txt",col\_names = T)

clusters$ID <- substr(rownames(clusters),1,15)

clusters$clusters <- as.character(clusters$clusters)

clusters%>%

left\_join(luad\_ssis)%>%

dplyr::select(-sample\_s,-ID)->ssis\_df

rownames(ssis\_df) <- rownames(clusters)

melt(ssis\_df)->ssis\_df1

ggplot(ssis\_df1,aes(x=clusters,y=value,fill=variable))+

geom\_boxplot()+

theme\_bw()->p

ggsave(filename = "estimate\_score/dist.luad.pdf",p,width = 11,height = 5)

#07-survival

#LUSC

clin\_lusc, dataSmNT,dataSmTP

expr,mirna\_exp\_clu

head(mirna\_exp\_clu)

rownames(mirna\_exp\_clu) <- mirna\_exp\_clu$miRNA\_ID

mirna\_exp\_clu%>%

dplyr::select(-miRNA\_ID)->mirna\_exp\_clu\_surdf

colnames(mirna\_exp\_clu\_surdf) <- substr(colnames(mirna\_exp\_clu\_surdf),32,46)

expr <- mirna\_exp\_clu\_surdf

intersect(rownames(expr),gene\_symbol)

#mutation sample

maf\_luad%>%

filter(Hugo\_Symbol %in% c("ATM","TP53"))%>%

dplyr::select(Tumor\_Sample\_Barcode)->mut\_sample\_df

mut\_sample<- intersect(substr(unique(mut\_sample\_df$Tumor\_Sample\_Barcode),1,15),colnames(expr))

sur<-function(gene\_name,gene\_symbol){

gene\_symbol <-"hsa-mir-185"

#outdir <- c("C:/Users/FuXin/Desktop/FU/PROJECT/Prostate\_Cancer/6.survival/Survival\_Month/chemokine")

outdir <- paste0("C:/Users/FuXin/OneDrive/FU/LiKe/LungCancer-202112-202206/analysis/08.survival")

setwd(outdir)

gene\_exp\_TP\_marker<-expr[gene\_symbol,intersect(substr(dataSmTP,1,15),mut\_sample)]

#gene\_exp\_TP\_marker<-expr[gene\_symbol,intersect(substr(dataSmTP,1,15),colnames(expr))]

names(gene\_exp\_TP\_marker) <- sapply(strsplit(names(gene\_exp\_TP\_marker),'-'),function(x) paste0(x[1:3],collapse="-"))

gene\_exp\_TP\_marker\_t <- t(gene\_exp\_TP\_marker)

clin\_luad$marker <- gene\_exp\_TP\_marker\_t[match(clin\_luad$submitter\_id,rownames(gene\_exp\_TP\_marker\_t)),]

clin\_luad%>%

filter(!is.na(marker))->clin\_luad.filt

df<-subset(clin\_luad.filt,

select =c(submitter\_id,vital\_status,days\_to\_death,days\_to\_last\_follow\_up,marker))

df$days\_to\_death[is.na(df$days\_to\_death)] <- 0

df$days\_to\_last\_follow\_up[is.na(df$days\_to\_last\_follow\_up)] <- 0

df$days=as.numeric(df$days\_to\_death)+as.numeric(df$days\_to\_last\_follow\_up)

df$days\_to\_death=round(df$days/30,2)

df$exp <- ''

if(median(df$marker)==0){

print("Undo")

}else{

df[df$marker >= median(df$marker),]$exp <- "H"

df[df$marker < median(df$marker),]$exp <- "L"

TCGAanalyze\_survival(df,

clusterCol="exp",

legend = gene\_symbol,

risk.table = FALSE,

conf.int = FALSE,

color = c("Dark2"),

#main = gene\_symbol,

xlab = "Time since diagnosis (Monthes)",

dpi=1200,

height = 4, width = 5,

#dpi = 300,

filename = paste0(gene\_symbol,"\_survival.mut.pdf"))

}

}