**Supplementary File 1. R code script of the bioinformatical analysis.**

library(ggunchained)

library(ggplot2)

library(ggsignif)

library(tidyverse)

library(ggpubr)

library(Seurat)

#Figure1

projectID <- "GSE117570"

cancer <- "NSCLC"

GSE117570\_sample\_list <- c("GSM3304007\_P1\_Tumor","GSM3304008\_P1\_Normal","GSM3304009\_P2\_Tumor","GSM3304010\_P2\_Normal",

 "GSM3304011\_P3\_Tumor","GSM3304012\_P3\_Normal","GSM3304013\_P4\_Tumor","GSM3304014\_P4\_Normal")

for(i in 1:8){

 sample = GSE117570\_sample\_list[i]

 mtx\_file <- paste0(analysis\_path,"/1.data/GSE117570/",sample,"\_processed\_data.txt.gz")

 mtx <- read.table(mtx\_file)

 sobj <- CreateSeuratObject(counts = mtx, min.features = 100, min.cells = 10, project = sample)

 sobj[["percent.mito"]] <- PercentageFeatureSet(sobj, pattern = "^MT-")

 VlnPlot(sobj, features = c("nFeature\_RNA", "nCount\_RNA", "percent.mito"), ncol = 3)->p

 p\_file <- paste0(analysis\_path,"/2.qc/GSE117570/",sample,".QC.pdf")

 ggsave(p\_file,p)

 sobj <- subset(sobj, subset = nFeature\_RNA > 100 & nFeature\_RNA < 5000 & percent.mito < 20 & nCount\_RNA > 10)

 assign(paste0("GSE117570\_sobj",i),sobj)

}

sobj\_list <- c(GSE117570\_sobj1,GSE117570\_sobj2,GSE117570\_sobj3,GSE117570\_sobj4,GSE117570\_sobj5,GSE117570\_sobj6,GSE117570\_sobj7,GSE117570\_sobj8)

GSE117570\_merge\_sobj <- merge(x = GSE117570\_sobj1,

 y = c(GSE117570\_sobj2,GSE117570\_sobj3,GSE117570\_sobj4,GSE117570\_sobj5,GSE117570\_sobj6,GSE117570\_sobj7,GSE117570\_sobj8))

#GSE162498

cancer <- "NSCLC"

projectID <- "GSE162498"

sample\_dir\_list <- c("P34\_Tumor\_raw","P35\_Tumor\_raw","P42\_Tumor\_raw","P43\_Tumor\_raw","P46\_Tumor\_raw","P47\_Tumor\_raw","P55\_Tumor\_raw",

 "P57\_Blood\_raw","P57\_Tumor\_raw","P58\_Blood\_filtered","P58\_Tumor\_filtered","P60\_Blood\_filtered","P60\_Juxta\_ffiltered",

 "P60\_Tumor\_filtered","P61\_Blood\_filtered","P61\_Juxta\_filtered","P61\_Tumor\_filtered")

sample\_list <- c("P34\_Tumor","P35\_Tumor","P42\_Tumor","P43\_Tumor","P46\_Tumor","P47\_Tumor","P55\_Tumor",

 "P57\_Blood","P57\_Tumor","P58\_Blood","P58\_Tumor","P60\_Blood","P60\_Juxta",

 "P60\_Tumor","P61\_Blood","P61\_Juxta","P61\_Tumor")

for(i in 1:17){

 print(i)

 mtx\_file <- paste0(analysis\_path,"/1.data/GSE162498/",sample\_dir\_list[i],"\_feature\_bc\_matrix")

 mtx <- Read10X(mtx\_file)

 sobj <- CreateSeuratObject(counts = mtx, min.features = 100, min.cells = 10, project = sample\_list[i])

 sobj[["percent.mito"]] <- PercentageFeatureSet(sobj, pattern = "^MT-")

 VlnPlot(sobj, features = c("nFeature\_RNA", "nCount\_RNA", "percent.mito"), ncol = 3)->p

 p\_file <- paste0(analysis\_path,"/2.qc/",projectID,"/",sample\_list[i],".QC.pdf")

 ggsave(p\_file,p)

 sobj <- subset(sobj, subset = nFeature\_RNA > 100 & nFeature\_RNA < 5000 & percent.mito < 20 & nCount\_RNA > 10)

 assign(paste0("Sobj",i),sobj)

}

sobj\_list <- c( Sobj1,Sobj2, Sobj3, Sobj4, Sobj5, Sobj6, Sobj7, Sobj8, Sobj9, Sobj10, Sobj11, Sobj12, Sobj13, Sobj14, Sobj15,

 Sobj16, Sobj17)

GSE162498\_merge\_sobj <- merge(x = Sobj1,y = c(Sobj2, Sobj3, Sobj4, Sobj5, Sobj6, Sobj7, Sobj8, Sobj9, Sobj10, Sobj11, Sobj12, Sobj13, Sobj14, Sobj15,

 Sobj16, Sobj17))

#remove batch

cancer <- "NSCLC"

project\_list <- c("GSE117570","GSE162498")

P\_rds\_1 <- GSE117570\_merge\_sobj

P\_rds\_2 <- GSE162498\_merge\_sobj

sobj\_list <- c(P\_rds\_1,P\_rds\_2)

sobj\_list <- lapply(X = sobj\_list, FUN = function(x) {

 x <- NormalizeData(x)

 x <- FindVariableFeatures(x, selection.method = "vst", nfeatures = 2000)

})

features <- SelectIntegrationFeatures(object.list = sobj\_list)

sobj\_anchors <- FindIntegrationAnchors(object.list = sobj\_list, anchor.features = features)

sobj\_combined <- IntegrateData(anchorset = sobj\_anchors)

rds <- sobj\_combined

use.genes <- rds@assays$integrated@var.features

rds <- ScaleData(rds, features = use.genes)

rds <- RunPCA(object = rds, features = use.genes, do.print = FALSE)

rds <- FindClusters(rds, resolution = 0.6)

rds = RunUMAP(rds, dims=1:20)

rds <- RunTSNE(rds, dims = 1:20, do.fast = TRUE, check\_duplicates = FALSE)

#Figure1.A,B,C

DimPlot(rds, reduction = "umap",label = TRUE)->p

p\_file <- paste0(analysis\_path,"/3.cluster/umap.cluster.pdf")

ggsave(p\_file,p,width = 8)

DimPlot(rds, reduction = "umap",group.by = "sample")->p

p\_file <- paste0(analysis\_path,"/3.cluster/umap.sample.pdf")

ggsave(p\_file,p,width = 8)

DimPlot(rds, reduction = "umap",group.by = "sample",split.by = "sample")->p

p\_file <- paste0(analysis\_path,"/3.cluster/umap.sample.split.pdf")

ggsave(p\_file,p,width = 8)

#Celltype annotation(see Method)

nsclc\_celltype <- list('0' = "Macrophage" ,

 '1' = "Th cell(CD4+)",

 '2' = "Cytotoxic T cell(CD8+)",

 '3' = "Cytotoxic T cell(CD8+)",

 '4' = "Th cell(CD4+)",

 '5' = "T cell(CD4+CD8+)",

 '6' = "Treg cell(CD4+)",

 '7' = "Th cell(CD4+)",

 '8' = "Naiive T cell(CD4-CD8-)",

 '9' = "Macrophage",

 '10' = "Cytotoxic T cell(CD8+)",

 '11' = "Naiive T cell(CD4-CD8-)",

 '12' = "Th cell(CD4+)",

 '13' = "Macrophage:M2",

 '14' = "Macrophage",

 '15' = "Monocyte",

 '16' = "T cell(CD4+CD8+)",

 '17' = "Monocyte",

 '18' = "Cancer cell",

 '19' = "Epithelial cell",

 '20' = "Cancer cell",

 '21' = "B cell",

 '22' = "Cancer cell",

 '23' = "Epithelial cell",

 '24' = "Macrophage",

 '25' = "Endothelial cell",

 '26' = "Epithelial cell",

 '27' = "Epithelial cell",

 '28' = "Naiive T cell(CD4-CD8-)")

rds\_file <- "5.celltype\_annotation/celltype.rds"

rds = readRDS(rds\_file)

#Figure1.D

celltype\_color = c("DarkOrange","GreenYellow","Purple","DarkSlateGray","Gold","DeepPink2","Red4","#4682B4",

 "#FFDAB9","#708090","#836FFF","#CDC673","#CD9B1D","#FF6EB4","#CDB5CD","DarkGreen")

names(celltype\_color)<-unique(rds$celltype)

DimPlot(rds,group.by = "celltype",cols = celltype\_color,label = TRUE, repel =TRUE)->p

p\_file <- paste0("5.celltype\_annotation/celltype.umap.pdf")

ggsave(p\_file,p,width = 8.5)

#Figure1.E,F,G

ggplot(rds@meta\_data,aes(x = sample\_class,fill = celltype))+

geom\_bar(width = 0.8, position = "fill")+

theme\_classic()+

scale\_fill\_manual(values = celltype\_color)+

theme(

 axis.text.x=element\_text(color = "black", size=15,angle = 30,hjust =1),

 axis.text.y=element\_text(color = "black",size=15),

 axis.title.x=element\_text(color = "black", size = 15),

 axis.title.y=element\_text(color = "black", size = 15),

 legend.text = element\_text(color = "black", size=15))->p.f1.E

ggplot(rds@meta\_data%>%filter(celltype %in% c("Cytotoxic T cell(CD8+)","Th cell(CD4+)",

 "T cell(CD4+CD8+)","Treg cell(CD4+)","Naiive T cell(CD4-CD8-)")),

 aes(x = sample\_class,fill = celltype))+

geom\_bar(width = 0.8, position = "fill")+

theme\_classic()+

scale\_fill\_manual(values = celltype\_color)+

theme(

 axis.text.x=element\_text(color = "black", size=15,angle = 30,hjust =1),

 axis.text.y=element\_text(color = "black",size=15),

 axis.title.x=element\_text(color = "black", size = 15),

 axis.title.y=element\_text(color = "black", size = 15),

 legend.text = element\_text(color = "black", size=15))->p.f1.F

ggplot(rds@meta\_data%>%filter(celltype %in% c("Macrophage:M2","Macrophage","Monocyte")),

 aes(x = sample\_class,fill = celltype))+

geom\_bar(width = 0.8, position = "fill")+

theme\_classic()+

scale\_fill\_manual(values = celltype\_color)+

theme(

 axis.text.x=element\_text(color = "black", size=15,angle = 30,hjust =1),

 axis.text.y=element\_text(color = "black",size=15),

 axis.title.x=element\_text(color = "black", size = 15),

 axis.title.y=element\_text(color = "black", size = 15),

 legend.text = element\_text(color = "black", size=15))->p.f1.G

#Figure2

rds\_file <- "5.celltype\_annotation/celltype.rds"

rds = readRDS(rds\_file)

marker\_cluster = FindAllMarkers(rds)

marker\_cluster%>%filter(p.val<0.05)%>%

 group\_by(cluster)%>%

 top\_n(10,wt=avg\_log2FC)->top10\_degs

DoHeatmap(rds,features = unique(top10\_degs$gene))

for(gene in unique(top10\_degs$gene)){

 FeaturePlot(rds,features = gene)->p

 p\_file = paste0("5.celltype\_annotation/fp.",gene,".pdf")

 ggsave(p\_file,p)

}

library(clusterProfiler)

library(org.Hs.eg.db)

GO\_enrich <- function(genelist,cluster){

 eg <- bitr(genelist,

 fromType="SYMBOL",

 toType=c("ENTREZID","ENSEMBL",'SYMBOL'),

 OrgDb="org.Hs.eg.db")

 go <- enrichGO(eg$ENTREZID,

 OrgDb = org.Hs.eg.db,

 ont='ALL',

 pAdjustMethod = 'BH',

 pvalueCutoff = 0.05,

 qvalueCutoff = 0.05,

 keyType = 'ENTREZID',

 readable = TRUE)

 write.table(go,,filename = paste0("go.",cluster,".csv"), sep=",")

 dotplot(go,showCategory=25,orderBy = "p.adjust")->p

 ggsave(paste0("go.",cluster,".pdf"),p)

 return(go)

}

KEGG\_enrich <- function(genelist,cluster){

 eg <- bitr(genelist,

 fromType="SYMBOL",

 toType=c("ENTREZID","ENSEMBL",'SYMBOL'),

 OrgDb="org.Hs.eg.db")

 kegg <- enrichKEGG(eg$ENTREZID,

 organism = 'hsa',

 keyType = 'kegg',

 pvalueCutoff = 0.05,

 pAdjustMethod = 'BH',

 minGSSize = 3,

 maxGSSize = 500,

 qvalueCutoff = 0.05,

 use\_internal\_data = FALSE)

 kegg <- setReadable(kegg, OrgDb = org.Hs.eg.db, keyType="ENTREZID")

 write.table(kegg,,filename = paste0("kegg.",cluster,".csv"), sep=",")

 barplot(kegg,showCategory=25,orderBy = "p.adjust")->p

 ggsave(paste0("kegg.",cluster,".pdf"),p)

 return(kegg)

}

marker\_cluster%>%filter(p.val<0.05)->marker\_cluster\_flt

for(clu in unique(marker\_cluster\_flt$cluster)){

 marker\_cluster\_flt%>%filter(cluster==clu)->tmp\_df

 genelist = unique(tmp\_df$gene)

 GO\_enrich(genelist,clu)

 KEGG\_enrich(genelist,clu)

}

#Figrue3

rds\_file <- "5.celltype\_annotation/celltype.rds"

rds = readRDS(rds\_file)

Idents(rds) <- "celltype"

FindAllmarkers(rds)->marker\_celltype

for(ct in unique(marker\_celltype$cluster)){

 subset(rds, celltype == ct) -> rds\_ct

 # Normal Vs Tumor

 diff\_dat\_N\_T <- FindMarkers(rds\_ct,ident.1="Normal",ident.2="Tumor",

 group.by='sample\_class')

 colnames(diff\_dat\_N\_T) <- paste0(colnames(diff\_dat\_N\_T),"\_N\_T")

 diff\_dat\_N\_T%>%

 mutate(gene = rownames(diff\_dat\_N\_T))%>%

 filter(p\_val\_adj\_N\_T < 0.001)->diff\_dat\_N\_T

 diff\_dat\_N\_T%>%

 select(avg\_log2FC\_N\_T,p\_val\_adj\_N\_T,gene)%>%

 mutate(Group = "N.vs.T")->diff\_dat\_N\_T\_1

 colnames(diff\_dat\_N\_T\_1) <- c("log2FC","P.val.adj","Gene","Group")

 # Normal Vs PBMC

 diff\_dat\_N\_B <- FindMarkers(rds\_ct,ident.1="Normal",ident.2="Blood",

 group.by='sample\_class')

 colnames(diff\_dat\_N\_B) <- paste0(colnames(diff\_dat\_N\_B),"\_N\_B")

 diff\_dat\_N\_B%>%

 mutate(gene = rownames(diff\_dat\_N\_B))%>%

 filter(p\_val\_adj\_N\_B < 0.001)->diff\_dat\_N\_B

 diff\_dat\_N\_B%>%

 select(avg\_log2FC\_N\_B,p\_val\_adj\_N\_B,gene)%>%

 mutate(Group = "N.vs.B")->diff\_dat\_N\_B\_1

 colnames(diff\_dat\_N\_P\_1) <- c("log2FC","P.val.adj","Gene","Group")

 #Tumor Vs Blood

 diff\_dat\_T\_B <- FindMarkers(rds\_ct,ident.1="Tumor",ident.2="Blood",

 group.by='sample\_class')

 colnames(diff\_dat\_T\_B) <- paste0(colnames(diff\_dat\_T\_B),"\_T\_B")

 diff\_dat\_T\_B%>%

 mutate(gene = rownames(diff\_dat\_T\_B))%>%

 filter(p\_val\_adj\_T\_B < 0.001)->diff\_dat\_T\_B

 diff\_dat\_T\_B%>%

 select(avg\_log2FC\_T\_B,p\_val\_adj\_T\_B,gene)%>%

 mutate(Group = "T.vs.B")->diff\_dat\_T\_B\_1

 colnames(diff\_dat\_T\_B\_1) <- c("log2FC","P.val.adj","Gene","Group")

 diff\_data<-rbind(diff\_dat\_N\_T\_1,diff\_dat\_N\_B\_1,diff\_dat\_T\_B\_1)

 marker\_celltype%>%filter(cluster==ct,p.val<0.05)->marker\_celltype\_spe

 diff\_data%>%filter(Gene %in% marker\_celltype\_sep$gene)->diff\_data\_spe

 write.table(diff\_data\_spe,"filename")

 ggplot(diff\_data\_spe,aes(x = Group,y = Gene))+

 geom\_point(aes(fill = log2FC,size = -log10(`P.val.adj`)),shape = 21,alpha = 0.8)+

 scale\_fill\_continuous( high=clustcol[17])+

 theme\_classic()+

 theme(

 axis.text.x=element\_text(color = "black", size=12),

 axis.text.y=element\_text(color = "black", size=7),

 axis.title.x=element\_text(color = "black", size = 13,face = "bold"),

 axis.title.y=element\_text(color = "black", size = 13,face = "bold")

 )->p

 p\_file <- paste0("5.celltype\_annotation/marker\_diff\_exp/",ct,".pdf")

 ggsave(p\_file,p,height = 8,width = 6.5)

}

#Survival analysis

library(survminer)

#exp = readr::read\_tsv("reference/TCGA/TCGA-LUAD.htseq\_fpkm.tsv.gz")

#sur\_df = readr::read\_tsv("reference/TCGA/TCGA-LUAD.survival.tsv")

exp = readr::read\_tsv("reference/TCGA/TCGA-LUSC.htseq\_counts.tsv.gz")

sur\_df = readr::read\_tsv("reference/TCGA/TCGA-LUSC.survival.tsv")

k=keys(org.Hs.eg.db,keytype = "ENSEMBL")

glist=select(org.Hs.eg.db,keys=k,columns = c("ENTREZID","SYMBOL"), keytype="ENSEMBL")

data.frame(gene=character(),pvalue = double(),celltype = character(),group = character())->pvalue\_df

for(ct in unique(rds$celltype)){

 diff\_dat\_spe\_file <- paste0("5.celltype\_annotation/marker\_diff\_exp/",ct,".diff.spe.csv")

 diff\_dat\_spe <- readr::read\_csv(diff\_dat\_spe\_file)

 for(i in 1:nrow(diff\_dat\_spe)){

 gene = diff\_dat\_spe$gene\_symbol[i]

 group = diff\_dat\_spe$group[i]

 print(gene)

 glist%>%

 filter(SYMBOL == gene)-> gene\_trans

 gene\_trans[order(gene\_trans$ENSEMBL),]->gene\_trans

 if(nrow(gene\_trans)>0){

 for(ensembl in gene\_trans$ENSEMBL){

 data.frame(t(exp[grepl(ensembl,exp$Ensembl\_ID),-1]))->gene\_exp

 if(ncol(gene\_exp)>0){break}

 }

 colnames(gene\_exp) <- c("exp")

 if(nrow(gene\_exp)>1&sum(gene\_exp$exp)>0){

 data.frame(gene\_exp)%>%mutate(sample = rownames(gene\_exp))->gene\_exp

 sur\_df%>%

 left\_join(gene\_exp)->sur\_df\_exp

 #print(head(sur\_df\_exp))

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

 sur\_df\_exp%>%

 group\_by(OS,`\_PATIENT`,`OS.time`)%>%

 summarize(exp\_fit = mean(exp))%>%

 ungroup()%>%

 mutate(group = ifelse(exp\_fit > median(exp\_fit),"High","Low"))%>%

 dplyr::select(-exp\_fit)->sur\_df\_exp

 colnames(sur\_df\_exp) <- c("Status","Patient","Time","Group")

 fit\_1 <- survfit(Surv(Time, Status) ~ Group, #group factor: sex

 data = sur\_df\_exp)

 #print(head(sur\_df\_exp))

 #print(fit)

 surv\_pvalue(fit\_1)->sp

 data.frame(gene = gene, pvalue = sp$pval,celltype = ct,group = group)->tmp\_df

 pvalue\_df <- rbind(pvalue\_df,tmp\_df)

 #print(pvalue\_df)

 ggsurvplot(fit\_1,

 pval = TRUE,

 conf.int = TRUE,

 risk.table = TRUE, # Add risk table

 risk.table.col = "Group", # Change risk table color by groups

 #linetype = "Group", # Change line type by groups

 surv.median.line = "hv", # Specify median survival

 ggtheme = theme\_bw(), # Change ggplot2 theme

 palette = c("#E7B800", "#2E9FDF"))->p

 pdf(paste0("08.survival.features/LUSC/",ct,".",gene,".pdf"))

 print(p)

 dev.off()

 }else{print("no exp")}

 }else{print("No gene found")}

 }

}

#Figure4

pvalue\_df%>%filter(pvalue<0.05)-pamg\_df

for(gene in pamg\_df$gene){

 FeaturePlot(rds,feature = gene)->p

 p\_file = paste0("6.marker.diff.spe/fp.",gene,".pdf")

 ggsave(p\_file,p)

}

DoHeatmap(rds,features = pamg\_df$gene,gruop.by = "celltype")

#plot KRT6A,ADM,NAPSA

VlnPlot(rds,featrues = "KRT6A",group.by = "celltype",split.by = "sample\_class")

rds[["KRT6A"]]<-rds@assays$RNA@counts["KRT6A",]

options(repr.plot.height=4 , repr.plot.width=12)

ggplot(gene\_meta%>%filter(sample\_class %in% c("Blood","Normal"),KRT6A>0),aes(x = celltype,y = log2(KRT6A+1),fill = sample\_class))+

geom\_split\_violin()+

theme\_classic()+

scale\_fill\_manual(values = list("Normal"="#0066CC","Blood"="#FFFF00"))+ #"Tumor" = "#FF6666","Blood"="#FFFF00","Normal"="#0066CC"

ylab("")+

xlab("")+

stat\_compare\_means(aes(group = sample\_class),label = "p.signif",hide.ns = TRUE)+

theme(

 axis.text.y = element\_text(color = "black", size=15), #不显示坐标刻度

 #axis.ticks.y = element\_text(color = "black", size=15),

 axis.title = element\_text(color = "black", size=13),

 axis.text.x = element\_text(color = 'black',size = 13,angle = 30,hjust =1,vjust = 1),

 legend.title = element\_text(color = "black", size=13),

 legend.text = element\_text(color = "black", size=13)

)->p

p

 #legend.position = 'none')

p\_file = paste0(analysis\_path,"/vln.B.N.KRT6A.gt0.pdf")

ggsave(p\_file,p,width = 11,height = 4.5)

#Figure5

marker\_celltype%>%filter(p.val<0.05)->marker\_celltype\_flt

for(clu in unique(marker\_celltype\_flt$cluster)){

 marker\_cluster\_flt%>%filter(cluster==clu)->tmp\_df

 genelist = unique(tmp\_df$gene)

 GO\_enrich(genelist,clu)

 KEGG\_enrich(genelist,clu)

}

#Figrue6

library(stringr)

library(tidyverse)

library(enrichplot)

hall\_gmt<-read.gmt("/Personal/fuxin/dfuxin/Bioinfo\_Scrs/scRNA\_biological\_analysis/reference/MigDB/h.all.v7.5.1.entrez.gmt")

immun\_gmt <- read.gmt("/Personal/fuxin/dfuxin/Bioinfo\_Scrs/scRNA\_biological\_analysis/reference/MigDB/c7.all.v7.5.1.entrez.gmt")

rds\_file <- "5.celltype\_annotation/celltype.rds"

obj <- readRDS(rds\_file)

diff\_dat\_sum = readr::read\_csv("6.marker\_diff\_spe/diff\_dat\_spe\_sum.csv")

for(ct in unique(rds$celltype)){

 diff\_dat\_sum%>%

 filter(celltype== ct) -> dds\_tmp

 genelist = unique(dds\_tmp$Gene)

 genelist = bitr(genelist,fromType="SYMBOL",toType="ENTREZID",OrgDb="org.Hs.eg.db")

 genelist <- dplyr::distinct(genelist,SYMBOL,.keep\_all=TRUE)

 genelist%>%

 left\_join(dds\_tmp, by = c("SYMBOL" = "Gene"))->genelist

 genelist[order(genelist$log2FC,decreasing = TRUE),]->genelist

 geneList <- genelist$log2FC

 names(geneList) <- genelist$ENTREZID

 hall\_res<-GSEA(geneList,TERM2GENE = hall\_gmt)

 immun\_res <- GSEA(geneList,TERM2GENE = immun\_gmt)

 if(nrow(immun\_res)>0){

 if(nrow(immun\_res)==1){

 write.table(immun\_res,paste0(gsea\_dir,"/",ct,".immun.csv"))

 gseaplot2(immun\_res, geneSetID = 1, title = immun\_res$Description[1])->p

 p\_file <- paste0(gsea\_dir,"/",ct,".immun.gsea.pdf")

 ggsave(p\_file,p,width = 6,height = 5)

 }else{

 write.table(immun\_res,paste0(gsea\_dir,"/",ct,".immun.csv"))

 data.frame(immun\_res)->immun\_res\_df

 immun\_res\_df[order(immun\_res\_df$p.adjust),]->immun\_res\_df

 ggplot(immun\_res\_df[1:20,], aes(NES, fct\_reorder(Description, NES), fill=qvalues)) +

 geom\_bar(stat='identity',width = 0.8) +

 scale\_fill\_continuous(high='#99CC66', low='#FF6666', guide=guide\_colorbar(reverse=TRUE)) +

 theme\_bw() +

 theme(

 axis.text.x=element\_text(color = "black", size=13),

 axis.text.y=element\_text(color = "black", size=13),

 axis.title.x=element\_text(color = "black", size = 13 ,face = "bold"),

 axis.title.y=element\_text(color = "black", size = 13,face = "bold")

 )+ylab(NULL)->p

 p\_file <- paste0(gsea\_dir,"/",ct,".immun.bar.pdf")

 ggsave(p\_file,p, width = 16,height = 8)

 gseaplot2(immun\_res, geneSetID = 1,title = immun\_res$Description[1])->p

 p\_file <- paste0(gsea\_dir,"/",ct,".immun.gsea.pdf")

 ggsave(p\_file,p,width = 6,height = 5)

 dotplot(immun\_res,split=".sign")+facet\_grid(~.sign)->p

 p\_file <- paste0(gsea\_dir,"/",ct,".immun.dot.pdf")

 ggsave(p\_file,p, width = 16,height = 8)

 }

 }

 if(nrow(hall\_res)>0){

 if(nrow(hall\_res)==1){

 write.table(immun\_res,paste0(gsea\_dir,"/",ct,".hall.csv"))

 gseaplot2(hall\_res, geneSetID = 1, title = hall\_res$Description[1])->p

 p\_file <- paste0(gsea\_dir,"/",ct,".hall.gsea.pdf")

 ggsave(p\_file,p,width = 10,height = 5)

 }else{

 write.table(hall\_res,paste0(gsea\_dir,"/",ct,".hall.csv"))

 data.frame(hall\_res)->hall\_res\_df

 hall\_res\_df[order(hall\_res\_df$p.adjust),]->hall\_res\_df

 ggplot(hall\_res\_df[1:20,], aes(NES, fct\_reorder(Description, NES), fill=qvalues)) +

 geom\_bar(stat='identity',width = 0.8) +

 scale\_fill\_continuous(high='#99CC66', low='#FF6666', guide=guide\_colorbar(reverse=TRUE)) +

 theme\_bw() +

 theme(

 axis.text.x=element\_text(color = "black", size=12),

 axis.text.y=element\_text(color = "black", size=12),

 axis.title.x=element\_text(color = "black", size = 13,face = "bold"),

 axis.title.y=element\_text(color = "black", size = 13,face = "bold")

 )+ylab(NULL)->p

 p\_file <- paste0(gsea\_dir,"/",ct,".hall.bar.pdf")

 ggsave(p\_file,p, width = 16,height = 8)

 gseaplot2(hall\_res, geneSetID = 1,title = hall\_res$Description[1])->p

 p\_file <- paste0(gsea\_dir,"/",ct,".hall.gsea.pdf")

 ggsave(p\_file,p,width = 8,height = 8)

 dotplot(hall\_res,split=".sign")+facet\_grid(~.sign)->p

 p\_file <- paste0(gsea\_dir,"/",ct,".hall.dot.pdf")

 ggsave(p\_file,p, width = 16,height = 8)

 }

}

#Figrue7

library(SingleCellExperiment)

library(Seurat)

library(GSEABase)

library(escape)

library(reshape2)

#Collected T cell marker from public database

immune\_cell\_marker

glist <- GeneSetCollection(immune\_cell\_marker))

es\_score <- enrichIt(obj = rds, gene.sets = glist, groups = 1000, cores = 5)

cbind(rds@meta.data,es\_score)->es\_score

melt(es\_score)->es\_score\_melt

es\_score\_melt$sample\_class <- factor(es\_score\_melt$sample\_class, levels = c("Normal","Tumor","Blood"))

ggplot(es\_score\_melt%>%filter(sample\_class %in% c("Blood","Tumor")),aes(x = variable,y = value,fill = sample\_class))+

geom\_boxplot()+

scale\_fill\_manual(values = list("Blood"="#FFFF00","Tumor" = "#FF6666"))+ #"Tumor" = "#FF6666","Blood"="#FFFF00","Normal"="#0066CC"

stat\_compare\_means(label = "p.signif",hide.ns = TRUE)+

theme\_classic()+

theme(

 axis.text.y = element\_text(color = "black", size=15), #不显示坐标刻度

 #axis.ticks.y = element\_text(color = "black", size=15),

 axis.title = element\_text(color = "black", size=13),

 axis.text.x = element\_text(color = 'black',size = 13,angle = 30,hjust =1,vjust = 1),

 legend.title = element\_text(color = "black", size=13),

 legend.text = element\_text(color = "black", size=13)

)->p

p

p\_file = paste0(analysis\_path,"/es\_score.all.B.T.pdf")

ggsave(p\_file,p,width = 12,height =5)

ggplot(es\_score\_melt%>%filter(celltype == "Th cell(CD4+)"),aes(x = variable,y = value))+

geom\_boxplot(fill = "Red4")+

scale\_fill\_manual(values = list("Blood"="#FFFF00","Tumor" = "#FF6666"))+ #"Tumor" = "#FF6666","Blood"="#FFFF00","Normal"="#0066CC"

stat\_compare\_means(label = "p.signif",hide.ns = TRUE,ref.group = ".all.")+

theme\_classic()+

theme(

 axis.text.y = element\_text(color = "black", size=15), #不显示坐标刻度

 #axis.ticks.y = element\_text(color = "black", size=15),

 axis.title = element\_text(color = "black", size=13),

 axis.text.x = element\_text(color = 'black',size = 13,angle = 30,hjust =1,vjust = 1),

 legend.title = element\_text(color = "black", size=13),

 legend.text = element\_text(color = "black", size=13)

)->p

p

p\_file = paste0(analysis\_path,"/es\_score.Th\_cell\_CD4.pdf")

ggsave(p\_file,p,width = 12,height =5)

#Figrue8

library(monocle3)

t\_rds\_file <- "7.Trajectory/T\_cell/tcell.rds"

t\_rds <- readRDS(t\_rds\_file)

data.frame(id = rownames(t\_rds@assays$integrated@data),

 gene\_short\_name = rownames(t\_rds@assays$integrated@data),

 num\_cells\_expressed = rowSums(t\_rds@assays$integrated@data != 0))->gene\_annotation

cds <- new\_cell\_data\_set(t\_rds@assays$integrated@data,

 cell\_metadata = t\_rds@meta.data,

 gene\_metadata = gene\_annotation)

cds <- preprocess\_cds(cds, num\_dim = 50)

cds <- align\_cds(cds, alignment\_group = "projectID")

cds <- reduce\_dimension(cds)

plot\_cells(cds, label\_groups\_by\_cluster=FALSE,

 color\_cells\_by = "celltype",

 #group\_label\_size = 3,

 label\_cell\_groups = FALSE

 )+

theme\_classic()->p

p\_file <- "6.Trajectory/T\_cell/trajectory.pdf"

ggsave(p\_file,p,width = 9)

cds <- learn\_graph(cds)

plot\_cells(cds, color\_cells\_by = "celltype",

 label\_groups\_by\_cluster = FALSE,

 label\_leaves = TRUE,

 label\_branch\_points = TRUE,

 graph\_label\_size = 1.5)

cds <- order\_cells(cds,root\_pr\_nodes = c('Y\_1','Y\_1013','Y\_1036','Y\_121','Y\_680'))

plot\_cells(cds,

 color\_cells\_by = "pseudotime",

 label\_cell\_groups = FALSE,

 label\_leaves = FALSE,

 label\_branch\_point = FALSE,

 graph\_label\_size = 1.5)+theme\_classic()->p

p\_file <- "6.Trajectory/T\_cell/pseudotime.pdf"

ggsave(p\_file,p,width =7)

#specific gene

cds\_pr\_res <- graph\_test(cds, neighbor\_graph = "principal\_graph",cores = 4)

#Figrue9

#TCR analysis

sample\_list <-c("GSM4952970\_P47\_Tumor","GSM4952971\_P55\_Tumor","GSM4952972\_P57\_Tumor","GSM4952973\_P57\_Blood","GSM4952974\_P58\_Tumor",

 "GSM4952975\_P58\_Blood","GSM4952976\_P60\_Tumor","GSM4952977\_P60\_Juxta","GSM4952978\_P60\_Blood","GSM4952979\_P61\_Tumor",

 "GSM4952980\_P61\_Juxta","GSM4952981\_P61\_Blood")

sample\_list\_brief <-c("P47\_Tumor","P55\_Tumor","P57\_Tumor","P57\_Blood","P58\_Tumor",

 "P58\_Blood","P60\_Tumor","P60\_Juxta","P60\_Blood","P61\_Tumor",

 "P61\_Juxta","P61\_Blood")

tcr\_df <- data.frame()

nb=6

for(i in 1:12){

 tcr\_file <- paste0("1.data/GSE162499/",

 sample\_list[i],"\_filtered\_contig\_annotations.csv.gz")

 tcr\_df\_tmp <- readr::read\_csv(tcr\_file)%>%

 mutate(sample = sample\_list\_brief[i])%>%

 mutate(barcode\_corr = paste0(barcode,"\_",nb))

 rbind(tcr\_df,tcr\_df\_tmp) -> tcr\_df

 nb = nb+1

}

data.frame(cele\_id = Cells(rds),sample = rds$orig.ident)->rds\_ids

separate(rds\_ids,cele\_id,into = c("barcode","rank"),sep = "\_")->rds\_ids

tcr\_df%>%

 inner\_join(rds\_ids)->tcr\_df\_1

subset(rds, cells = intersect(tcr\_df\_1$cell\_id,Cells(rds)))->rds\_tcr

DimPlot(rds\_tcr,reduction = "umap", group.by = "celltype",label = TRUE, repel = TRUE,split.by="sample\_class")->p

p\_file <- "10.tcr/TCR.detect.umap.split.pdf"

ggsave(p\_file,p,width = 12)

data.frame(barcode = Cells(rds\_tcr),celltype = rds\_tcr$celltype,sample\_class = rds\_tcr$sample\_class)%>%

 separate(barcode,into = c("barcode","barcode\_fix"),sep = '\_')%>%

 left\_join(tcr\_df)->tcr\_anno

ggplot(tcr\_anno,aes(x = sample\_class,y = cell\_freq,fill = celltype))+

geom\_bar(stat="identity",stack = "dodge")+

scale\_fill\_manual(values = clustcol[1:15])+

theme\_classic()+

 theme(

 axis.text.x=element\_text(color = "black", size=12),

 axis.text.y=element\_text(color = "black", size=7),

 axis.title.x=element\_text(color = "black", size = 13,face = "bold"),

 axis.title.y=element\_text(color = "black", size = 13,face = "bold")

 )->p

p\_file <- "10.tcr/population.bar.pdf"

ggsave(p\_file,p)

ggplot(tcr\_anno\_brief,aes(x = celltype,fill = sample\_class))+

geom\_bar(position = position\_dodge())+

scale\_fill\_manual(values = c("#FFFF00","#FF0033"))+

ylab("Contigs number")+

theme\_classic() +

 theme(

 axis.text.x=element\_text(color = "black", size=12,angle = 30,hjust = 1),

 axis.text.y=element\_text(color = "black", size=12),

 axis.title.x=element\_text(color = "black", size = 13,face = "bold"),

 axis.title.y=element\_text(color = "black", size = 13,face = "bold")

 )->p

p\_file <- "10.tcr/tcr.contig.num.pdf"

ggsave(p\_file,p,width = 10,height =5)

ggplot(tcr\_anno\_brief,aes(x = celltype,y = length,fill = sample\_class))+

geom\_boxplot()+

scale\_fill\_manual(values = c("#FFFF00","#FF0033"))+

theme\_classic() +

 theme(

 axis.text.x=element\_text(color = "black", size=12,angle = 30,hjust = 1),

 axis.text.y=element\_text(color = "black", size=12),

 axis.title.x=element\_text(color = "black", size = 13,face = "bold"),

 axis.title.y=element\_text(color = "black", size = 13,face = "bold")

 )->p

p\_file <- "10.tcr/tcr.length.pdf"

ggsave(p\_file,p,width = 10,height =5)