Supplementary File 1. The R code used in this study

# scCode

library(Seurat)

library(tidyverse)

options(stringsAsFactors = F)

mycol <- c("#D20A13","#223D6C","#FFD121","#91612D","#08FCFC","#11AA4D","#58CDD9","#7A142C","#4300FF","#5D90BA","#431A3D","#088247","#6E568C","#E0367A","#D8D155","#64495D","#7CC767")

expression\_matrix <-Read10X("./")

seurat\_obj <- CreateSeuratObject(count = expression\_matrix,

min.cells = 3,

min.features = 200)

#

seurat\_obj[['percent.mt']] <- PercentageFeatureSet(seurat\_obj, pattern = "^MT-")

pdf("1.VlnPlot.pdf")

VlnPlot(seurat\_obj, features = c("nFeature\_RNA", "nCount\_RNA", "percent.mt"), ncol = 3)

dev.off()

seurat\_obj <- subset(seurat\_obj, subset = nFeature\_RNA > 200 & nFeature\_RNA < 5000 & percent.mt < 30)

#

seurat\_obj <- NormalizeData(seurat\_obj,

normalization.method = "LogNormalize",

scale.factor = 10000)

# HVG(highly variable features)

seurat\_obj <- FindVariableFeatures(seurat\_obj,selection.method = "vst",

nfeatures = 2000,num.bin = 20)

#

all.genes <- rownames(seurat\_obj)

seurat\_obj <- ScaleData(seurat\_obj, features = all.genes)

# PCA

seurat\_obj <- RunPCA(seurat\_obj,

features = NULL,

npcs = 100,

ndims.print = 1:5,

nfeatures.print = 5)

#

pdf("2.VizDimLoadings.plot.pdf")

VizDimLoadings(seurat\_obj, dims = 1:2, reduction = "pca")

dev.off()

pdf("3.DimPlot.pdf")

DimPlot(seurat\_obj, reduction = "pca")

dev.off()

pdf("4.DimHeatmap.pdf")

DimHeatmap(seurat\_obj, dims = 1, cells = 500, balanced = TRUE)

dev.off()

pdf("5.ElbowPlot.pdf")

ElbowPlot(seurat\_obj)

dev.off()

#

seurat\_obj <- FindNeighbors(seurat\_obj, dims = 1:50)

seurat\_obj <- FindClusters(seurat\_obj, resolution = 0.5)

#

levels(seurat\_obj)

# save(seurat\_obj,file = "seurat\_obj.rdata")

#

library(ggplot2)

seurat\_obj <- Run(seurat\_obj,dims = 1:50,n.neighbors=30,min.dist=0.3)

pdf("6.DimPlot\_cluster.pdf",width = 7,height = 6)

DimPlot(seurat\_obj,

reduction = "",

pt.size = 0.2,

label = TRUE) +

theme(panel.border = element\_blank(),axis.line = element\_line(colour = "black",size=1),

panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank())

dev.off()

save(seurat\_obj,file = "seurat\_obj.RData")

# Anno

seurat\_obj\_DimRe = readRDS("./TempData/seurat\_obj\_DimRe.rds")

Color = readRDS("./TempData/Color.rds")

Size = read.table("./TempData/Size.txt",sep = "\t",header = T,quote = "")

Width = read.table("./TempData/Width.txt",sep="\t",header = T)

Height = read.table("./TempData/Height.txt",sep="\t",header = T)

s\_meta <-seurat\_obj\_DimRe@meta.data

CellType = read.table("./TempData/CellType.txt",sep="\t",header = T)

if("CellType" %in% colnames(s\_meta)){

ColName = data.frame(ColName = colnames(s\_meta))

ColName\_1 = data.frame(ColName,Num = rownames(ColName))

NumCellType = as.numeric(ColName\_1[ColName\_1$ColName == "CellType",2])

colnames(s\_meta)[NumCellType] = "CellType"

}

s\_meta2 = merge(s\_meta,CellType,by.x="seurat\_clusters",by.y="Cluster",all.x = T,sort=F)

rownames(s\_meta2) <-s\_meta2$cells

s\_meta3 = s\_meta2[rownames(s\_meta),]

seurat\_obj\_DimRe@meta.data <- s\_meta3

names(seurat\_obj\_DimRe@meta.data)

cell2type <-data.frame(cells=pred.cells@rownames,pred.cells=pred.cells$labels)

meta <-seurat\_obj@meta.data

meta$cells <-rownames(meta)

meta2 <-merge(meta,cell2type,by = "cells",sort = F)

rownames(meta2) <-meta2$cells

seurat\_obj@meta.data <- meta2

names(seurat\_obj@meta.data)

Idents(seurat\_obj) <- "pred.cells"

mycol <- c("#D20A13","#223D6C","#FFD121","#91612D","#08FCFC","#11AA4D","#58CDD9","#7A142C","#4300FF","#5D90BA","#431A3D","#088247","#6E568C","#E0367A","#D8D155","#64495D","#7CC767")

pdf("6.DimPlot.pdf",width = 7,height = 6)

DimPlot(seurat\_obj,

reduction = "",

pt.size = 0.1,

label = T,

cols = colorRampPalette(mycol)(16)) +

ggtitle("RefAML") +

theme(panel.border = element\_blank(),axis.line = element\_line(colour = "black",size=1),

panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank(),

plot.title = element\_text(hjust = 0.5))

dev.off()

cellTypes <-data.frame(table(cellType=pred.cells$labels,cluster=seurat\_obj$RNA\_snn\_res.0.5))

cellTypes <-cellTypes[cellTypes$Freq>0,]

cellTypes <-spread(data=cellTypes,cluster,value = Freq)

cellTypes[is.na(cellTypes)]<-0

write.csv(cellTypes,file = "cellTypes.csv",row.names = F,quote = F)

s\_meta <-seurat\_obj@meta.data

s\_meta$cells <-rownames(s\_meta)

s\_meta$seurat\_clusters <-as.character(s\_meta$seurat\_clusters)

seurat\_obj@meta.data <- s\_meta

names(seurat\_obj@meta.data)

Idents(seurat\_obj) <- "seurat\_clusters"

stack\_color <-c("#1177BB","#FF7700","#229966","#DD2222","#AA44FF","#885544","#EE77CC","#BBBB66","#11BBCC","#AACCEE","#FFBB77","#87ED88","#FF9999","#CCBBDD","#CC9999","#FFBBDD","#DDDD88","#99DDEE","#AA4444","#886633")

pdf("6.DimPlot\_final.pdf",width = 8,height = 8)

DimPlot(seurat\_obj,

reduction = "",

pt.size = 0.1,

label = F,

cols = stack\_color ) +

ggtitle("DimPlot") +

theme(panel.border = element\_blank(),axis.line = element\_line(colour = "black",size=1), # 去除默认填充的灰色，并将x=0轴和y=0轴加粗显示(size=1)

panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank(),

plot.title = element\_text(hjust = 0.5)) +

coord\_fixed(1)

dev.off()

seurat\_obj\_final <- seurat\_obj

save(seurat\_obj\_final,file = "seurat\_obj\_final.rdata")

load("data\_stemness.rdata")

meta <-seurat\_obj\_final@meta.data

ans <-merge(meta,data\_stemness,by.x="cells",by.y="Sample")

ans <-ans[,c("seurat\_clusters","StemnessScore")]

medianStem <-ans %>% group\_by(seurat\_clusters) %>% summarise(median=median(StemnessScore))

#

library(dplyr)

final\_cluster <- levels(seurat\_obj\_final)

# all markers

markers\_all <- FindAllMarkers(seurat\_obj\_final, only.pos = FALSE, min.pct = 0.25, logfc.thresh = 0.25)

write.table(markers\_all, "markers\_all.txt",row.names = F,quote = F,sep="\t")

markers\_ROC\_all <- FindAllMarkers(seurat\_obj\_final, only.pos = FALSE, test.use = "roc", min.pct = 0.25, logfc.thresh = 0.25)

write.table(markers\_ROC\_all, "markers\_ROC\_all.txt",row.names = F,quote = F,sep="\t")

#

markers\_pos <- FindAllMarkers(seurat\_obj\_final, only.pos = TRUE, min.pct = 0.25, logfc.thresh = 0.25)

write.table(markers\_pos, "markers\_pos.txt",row.names = F,quote = F,sep="\t")

markers\_ROC\_pos <- FindAllMarkers(seurat\_obj\_final, only.pos = TRUE, test.use = "roc", min.pct = 0.25, logfc.thresh = 0.25)

write.table(markers\_ROC\_pos, "markers\_ROC\_pos.txt",row.names = F,quote = F,sep="\t")

#

fcolor <-c("#134479","#2E76B1","#9ECAE0","#E0ECF2","white","#F8EAE0","#F5BDA3","#E07964","#A1132B")

expression.color <- c("darkblue", "lightblue", "green", "yellow", "red")

min <-seurat\_obj\_final@assays$RNA@data["PTPRC",] %>% range %>% .[1]

max <-seurat\_obj\_final@assays$RNA@data["PTPRC",] %>% range %>% .[2] %>% round(.,2)

#fcolor <-c("#134479","#71ADD0","#8ABFD9","#D0E3ED","gray99","#F8EEE7","#F9E9DF","#FBDCCB","#F6BDA3","#F1A384","#E99076","#BC323A","#7A0824")

#fcolor <-c("#134479","#2E76B1","#9ECAE0","#F7F5F4","#F5BDA3","#E07964","#A1132B")

pdf("8.FeaturePlot.pdf",width = 6,height = 6)

FeaturePlot(seurat\_obj\_final,

features ="PTPRC",

pt.size = 0.2,

cols = c("gray90","red"),

label = FALSE,

label.size = 4) +

scale\_color\_gradientn(colors = expression.color)

dev.off()

#

adj\_DoHeatmap <-function(obj,features){

return(DoHeatmap(obj,

features = features,

size = 2,

hjust = 1.5,

angle = 0,

label = TRUE,

draw.lines = TRUE,

lines.width = 20,

group.bar.height = 0.04,

group.colors = colorRampPalette(mycol)(20)

) + NoLegend() +

scale\_fill\_gradient2(low = "#1C61A6", high = "#EBA28C", mid = "#273C65"))

}

pdf("9.DoHeatmap\_ROC\_Top1.pdf",width = 10,height = 10)

adj\_DoHeatmap(seurat\_obj\_final,Show\_markers$gene)

dev.off()

library(Seurat)

library(ggplot2)

library(monocle3)

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

load("seurat\_obj\_final.Rdata")

metadata <-seurat\_obj\_final@meta.data

cells1 <-metadata$cells

gene\_annotation <- as.data.frame(rownames(seurat\_obj\_final@reductions[["pca"]]@feature.loadings),

row.names = rownames(seurat\_obj\_final@reductions[["pca"]]@feature.loadings))

colnames(gene\_annotation) <- "gene\_short\_name"

cell\_metadata <- as.data.frame(cells1,

row.names = cells1)

colnames(cell\_metadata) <- "barcode"

New\_matrix <- seurat\_obj\_final@assays$RNA@counts

New\_matrix <- New\_matrix[rownames(seurat\_obj\_final@reductions[["pca"]]@feature.loadings),cells1]

expression\_matrix <- New\_matrix

cds\_from\_seurat <- new\_cell\_data\_set(expression\_matrix,

cell\_metadata = cell\_metadata,

gene\_metadata = gene\_annotation)

cds\_from\_seurat@metadata

# tmp.cds <- preprocess\_cds(cds\_from\_seurat,method = "PCA",num\_dim = 100)

# tmp.cds <- reduce\_dimension(tmp.cds,preprocess\_method = "PCA")

# tmp.cds <- cluster\_cells(tmp.cds)

# tmp.cds@clusters@listData$$partitions

cds\_from\_seurat@clusters@listData[[""]][["clusters"]] <- Idents(seurat\_obj\_final)

cata(cds\_from\_seurat)$celltype <- seurat\_obj\_final@meta.data[rownames(cata(cds\_from\_seurat)),"seurat\_clusters"]

# partitions

cds\_from\_seurat@clusters@listData[[""]][["partitions"]] <- "1"

#

cds\_from\_seurat@clusters@listData[[""]][["louvain\_res"]] <- "NA"

#

cds\_from\_seurat@int\_cata@listData$reducedDims[[""]] <-seurat\_obj\_final@reductions[[""]]@cell.embeddings[cells1,]

colnames(cds\_from\_seurat@int\_cata@listData$reducedDims[[""]]) <-c("V1","V2")

# order\_cells

#

cds\_from\_seurat@preprocess\_aux$gene\_loadings <- seurat\_obj\_final@reductions[["pca"]]@feature.loadings

cds\_from\_seurat@int\_cata@listData$reducedDims$PCA <- seurat\_obj\_final@reductions[["pca"]]@cell.embeddings[cells1,]

#

cds\_from\_seurat\_2d <- learn\_graph(cds\_from\_seurat)

get\_earliest\_principal\_node <- function(cds, time\_bin="Prog"){

cell\_ids <- which(cata(cds)[, "celltype"] == time\_bin)

closest\_vertex <-

cds@principal\_graph\_aux[[""]]$pr\_graph\_cell\_proj\_closest\_vertex

closest\_vertex <- as.matrix(closest\_vertex[colnames(cds), ])

root\_pr\_nodes <-

igraph::V(principal\_graph(cds)[[""]])$name[as.numeric(names

(which.max(table(closest\_vertex[cell\_ids,]))))]

root\_pr\_nodes

}

p\_node\_2d\_1 <-get\_earliest\_principal\_node(cds\_from\_seurat\_2d,time\_bin = "HSC")

p\_node\_2d\_2 <-get\_earliest\_principal\_node(cds\_from\_seurat\_2d,time\_bin = "earlyEry")

cds\_from\_seurat\_2d <- order\_cells(cds\_from\_seurat\_2d,root\_pr\_nodes=c(p\_node\_2d\_1,p\_node\_2d\_2))

pse <-as.data.frame(pseudotime(cds\_from\_seurat\_2d))

#pse$`pseudotime(cds\_from\_seurat\_2d)` <-ceiling(pse$`pseudotime(cds\_from\_seurat\_2d)`)

save(pse,file="MM\_pseudotime.rdata")

plot\_cells(cds\_from\_seurat\_2d,

color\_cells\_by = "cluster",

#cell\_size = 2,

alpha = 0.6,

label\_roots= FALSE,

label\_leaves=FALSE,

graph\_label\_size = 2,

trajectory\_graph\_color = NA,

trajectory\_graph\_segment\_size = 0.5 ,

label\_cell\_groups = FALSE,

group\_label\_size = 5,

label\_branch\_points=FALSE) +

# scale\_colour\_gradient2(low="#134479", high="#A1132B",mid = "#F7F5F4",midpoint = 15)

scale\_discrete\_manual("colour",values = colorRampPalette(mycol)(16))

ggsave("M1.plot\_cells\_pseudotime.pdf",width = 6.5,height = 6)

ggsave("M1.plot\_cells\_cluster.pdf",width = 6.5,height = 6)

ciliated\_cds\_pr\_test\_res <- graph\_test(cds\_from\_seurat\_2d, neighbor\_graph="principal\_graph")

save(ciliated\_cds\_pr\_test\_res,file = "ciliated\_cds\_pr\_test\_res.rdata")

pr\_deg\_ids <- row.names(subset(ciliated\_cds\_pr\_test\_res, q\_value < 0.01))

gene\_module\_df <- find\_gene\_modules(cds\_from\_seurat\_2d[pr\_deg\_ids,], resolution= 0.01)

write.table(gene\_module\_df,"gene\_module\_df.txt",row.names = F,quote = F,sep = "\t")