**Source code 1**

R code for single cell RNAseq analysis for Figure 7A-D, Sup Fig 3A-D, Sup Fig 4A-C

library(Seurat)

library(tidyverse)

library(ggplot2)

library(patchwork)

Healthy.1 <- ReadMtx(mtx = "GSM3660641\_SC14NOR\_matrix.mtx.gz",

 features = "GSM3660641\_SC14NOR\_genes.tsv.gz",

 cells = "GSM3660641\_SC14NOR\_barcodes.tsv.gz")

Healthy.2 <- ReadMtx(mtx = "GSM3660642\_SC31NOR\_matrix.mtx.gz",

 features = "GSM3660642\_SC31NOR\_genes.tsv.gz",

 cells = "GSM3660642\_SC31NOR\_barcodes.tsv.gz")

Healthy.3 <- ReadMtx(mtx = "GSM3660643\_SC31DNOR\_matrix.mtx.gz",

 features = "GSM3660643\_SC31DNOR\_genes.tsv.gz",

 cells = "GSM3660643\_SC31DNOR\_barcodes.tsv.gz")

Healthy.4 <- ReadMtx(mtx = "GSM3660644\_SC45NOR\_matrix.mtx.gz",

 features = "GSM3660644\_SC45NOR\_genes.tsv.gz",

 cells = "GSM3660644\_SC45NOR\_barcodes.tsv.gz")

Healthy.5 <- ReadMtx(mtx = "GSM3660645\_SC56NOR\_matrix.mtx.gz",

 features = "GSM3660645\_SC56NOR\_genes.tsv.gz",

 cells = "GSM3660645\_SC56NOR\_barcodes.tsv.gz")

Healthy.6 <- ReadMtx(mtx = "GSM3660646\_SC59NOR\_matrix.mtx.gz",

 features = "GSM3660646\_SC59NOR\_genes.tsv.gz",

 cells = "GSM3660646\_SC59NOR\_barcodes.tsv.gz")

Healthy.7 <- ReadMtx(mtx = "GSM3660647\_SC155NORLOW\_matrix.mtx.gz",

 features = "GSM3660647\_SC155NORLOW\_genes.tsv.gz",

 cells = "GSM3660647\_SC155NORLOW\_barcodes.tsv.gz")

Healthy.8 <- ReadMtx(mtx = "GSM3660648\_SC156NORUP\_matrix.mtx.gz",

 features = "GSM3660648\_SC156NORUP\_genes.tsv.gz",

 cells = "GSM3660648\_SC156NORUP\_barcodes.tsv.gz")

sdata.Healthy\_1 <- CreateSeuratObject(Healthy.1, project = "Healthy\_1")

sdata.Healthy\_2 <- CreateSeuratObject(Healthy.2, project = "Healthy\_2")

sdata.Healthy\_3 <- CreateSeuratObject(Healthy.3, project = "Healthy\_3")

sdata.Healthy\_4 <- CreateSeuratObject(Healthy.4, project = "Healthy\_4")

sdata.Healthy\_5 <- CreateSeuratObject(Healthy.5, project = "Healthy\_5")

sdata.Healthy\_6 <- CreateSeuratObject(Healthy.6, project = "Healthy\_6")

sdata.Healthy\_7 <- CreateSeuratObject(Healthy.7, project = "Healthy\_7")

sdata.Healthy\_8 <- CreateSeuratObject(Healthy.8, project = "Healthy\_8")

sdata.Healthy\_1$type = "LUNG"

sdata.Healthy\_2$type = "LUNG"

sdata.Healthy\_3$type = "LUNG"

sdata.Healthy\_4$type = "LUNG"

sdata.Healthy\_5$type = "LUNG"

sdata.Healthy\_6$type = "LUNG"

sdata.Healthy\_7$type = "LUNG"

sdata.Healthy\_8$type = "LUNG"

alldata <- merge(sdata.Healthy\_1, c(sdata.Healthy\_2, sdata.Healthy\_3,sdata.Healthy\_4, sdata.Healthy\_5,sdata.Healthy\_6,

 sdata.Healthy\_7,sdata.Healthy\_8),

 add.cell.ids = c("Healthy\_1", "Healthy\_2","Healthy\_3","Healthy\_4", "Healthy\_5","Healthy\_6",

 "Healthy\_7", "Healthy\_8"))

rm(Healthy.1, Healthy.2, Healthy.3,Healthy.4, Healthy.5, Healthy.6,Healthy.7, Healthy.8,

 sdata.Healthy\_1, sdata.Healthy\_2, sdata.Healthy\_3,sdata.Healthy\_4, sdata.Healthy\_5,sdata.Healthy\_6,

 sdata.Healthy\_7,sdata.Healthy\_8)

gc()

modified\_alldata <- alldata

modified\_alldata <- PercentageFeatureSet(modified\_alldata, "^MT-", col.name = "percent\_mito")

VlnPlot(modified\_alldata, features = c("nFeature\_RNA", "nCount\_RNA", "percent\_mito"), ncol = 3)

modified\_alldata <- subset(modified\_alldata, subset = nFeature\_RNA > 200 & nFeature\_RNA < 6000

 & nCount\_RNA > 1000

 & percent\_mito < 15)

modified\_alldata <- NormalizeData(modified\_alldata)

modified\_alldata <- FindVariableFeatures(modified\_alldata, selection.method = "vst", nfeatures = 2000, verbose = FALSE)

top2000 <- head(VariableFeatures(modified\_alldata), 2000)

modified\_alldata <- modified\_alldata[top2000]

modified\_alldata <- JoinLayers(modified\_alldata)

library(reticulate)

library(sceasy)

use\_python("PATH/python.exe")

sc <- import("scanpy", convert = FALSE)

scvi <- import("scvi", convert = FALSE)

modified\_alldata[["RNA"]] <- as(modified\_alldata[["RNA"]], "Assay")

adata <- convertFormat(modified\_alldata, from="seurat", to="anndata", main\_layer="counts", drop\_single\_values=FALSE)

scvi$model$SCVI$setup\_anndata(adata)

model = scvi$model$SCVI(adata, n\_latent = 10L)

model$train()

latent = model$get\_latent\_representation()

latent <- as.matrix(latent)

rownames(latent) = colnames(modified\_alldata)

modified\_alldata[["scvi"]] <- CreateDimReducObject(embeddings = latent, key = "scvi\_", assay = DefaultAssay(modified\_alldata))

modified\_alldata <- FindNeighbors(modified\_alldata, dims = 1:10, reduction = "scvi")

modified\_alldata <- FindClusters(modified\_alldata, resolution = 1.2, reduction = "scvi")

modified\_alldata <- RunUMAP(modified\_alldata, dims = 1:10, reduction = "scvi", n.components = 2)

modified\_alldata\_markers <- FindAllMarkers(modified\_alldata, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25)

write.csv(modified\_alldata\_markers, file = "Markers res1.2.csv")

p1 <- DimPlot(modified\_alldata, group.by = "orig.ident", reduction = "scvi", label.size = 4, pt.size = 1.5)

p1

p2 <- DimPlot(modified\_alldata, group.by = "seurat\_clusters", reduction = "scvi", label.size = 4, pt.size = 1.5)

p2

p3 <- VlnPlot(modified\_alldata, features = c("nFeature\_RNA", "nCount\_RNA", "percent\_mito"),

 group.by = "orig.ident", pt.size = 0)

p3

p4 <- FeaturePlot(modified\_alldata, features = c("CD163","FABP4","LYVE1","FCN1","CD1C","VWF","CD3D","CD8A",

 "IGHM","KLRD1","TPSAB1","SFTPC","AGER","FOXJ1","MUC5AC","COL1A1"), reduction = "scvi")

p4

clusters\_to\_merge <- c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,

 19, 20, 21, 22, 23, 24, 25,26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36)

modified\_alldata$custom\_cluster <- Idents(modified\_alldata)

new\_cluster\_label\_1 <- "Airway Goblet Cells"

new\_cluster\_label\_2 <- "Alveolar Epithelial Type 1"

new\_cluster\_label\_3 <- "Ciliated Cell Lung"

new\_cluster\_label\_4 <- "DC"

new\_cluster\_label\_5 <- "Endothelial Cell Lung"

new\_cluster\_label\_6 <- "Lymphocyte"

new\_cluster\_label\_7 <- "Macrophage"

new\_cluster\_label\_8 <- "Mast Cell"

new\_cluster\_label\_9 <- "Monocyte"

new\_cluster\_label\_10 <- "NK"

new\_cluster\_label\_11 <- "Fibroblast"

new\_cluster\_label\_12 <- "Pneumocyte Lung Type II"

modified\_alldata$custom\_cluster <- factor(

 modified\_alldata$custom\_cluster,

 levels = c(levels(modified\_alldata$custom\_cluster), new\_cluster\_label\_1, new\_cluster\_label\_2,

 new\_cluster\_label\_3, new\_cluster\_label\_4,new\_cluster\_label\_5, new\_cluster\_label\_6,

 new\_cluster\_label\_7, new\_cluster\_label\_8,new\_cluster\_label\_9, new\_cluster\_label\_10,

 new\_cluster\_label\_11, new\_cluster\_label\_12))

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(26)] <- new\_cluster\_label\_1

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(31)] <- new\_cluster\_label\_2

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(17)] <- new\_cluster\_label\_3

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(25)] <- new\_cluster\_label\_4

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(15, 20, 21, 33)] <- new\_cluster\_label\_5

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(2, 10, 12, 30, 36)] <- new\_cluster\_label\_6

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(0, 1, 3, 5, 7, 8, 11, 16, 19, 23, 27, 28, 32, 35)] <- new\_cluster\_label\_7

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(24)] <- new\_cluster\_label\_8

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(4, 6, 29)] <- new\_cluster\_label\_9

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(9, 22)] <- new\_cluster\_label\_10

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(14, 34)] <- new\_cluster\_label\_11

modified\_alldata$custom\_cluster[modified\_alldata$custom\_cluster %in% c(13, 18)] <- new\_cluster\_label\_12

DimPlot(modified\_alldata, group.by = "custom\_cluster", reduction = "scvi", label = T,

 label.size = 2.5, repel = T, pt.size = 1.5)

p5 <- DimPlot(modified\_alldata, group.by = "seurat\_clusters", reduction = "scvi", label.size = 4, pt.size = 1.5)

p5

### Macrophage reintegration

Macrophage = subset(modified\_alldata,idents = c('0','1','3','4','5','6','7','8','11','16','19','23','27','28','29',

 '32','35'))

DefaultAssay(Macrophage) <- "RNA"

Macrophage <- NormalizeData(Macrophage)

Macrophage <- FindVariableFeatures(Macrophage, selection.method = "vst",nfeatures = 2000, verbose = FALSE)

top2000 <- head(VariableFeatures(Macrophage), 2000)

Macrophage <- Macrophage[top2000]

sc <- import("scanpy", convert = FALSE)

scvi <- import("scvi", convert = FALSE)

Macrophage[["RNA"]] <- as(Macrophage[["RNA"]], "Assay")

adata\_macrophage <- convertFormat(Macrophage, from = "seurat", to = "anndata", main\_layer = "counts", drop\_single\_values = FALSE)

scvi$model$SCVI$setup\_anndata(adata\_macrophage)

model\_macrophage <- scvi$model$SCVI(adata\_macrophage, n\_latent = 10L)

model\_macrophage$train()

latent\_macrophage <- model\_macrophage$get\_latent\_representation()

latent\_macrophage <- as.matrix(latent\_macrophage)

rownames(latent\_macrophage) <- colnames(Macrophage)

Macrophage[["scvi"]] <- CreateDimReducObject(embeddings = latent\_macrophage, key = "scvi\_", assay = DefaultAssay(Macrophage))

Macrophage <- FindNeighbors(Macrophage, dims = 1:10, reduction = "scvi")

Macrophage <- FindClusters(Macrophage, resolution = 0.6)

Macrophage <- RunUMAP(Macrophage, dims = 1:10, reduction = "scvi", n.components = 2)

pm1 <- DimPlot(Macrophage, reduction = "scvi", label = T)

pm1

pm2 <- DimPlot(Macrophage, reduction = "scvi", group.by = "orig.ident", label = F, pt.size = 1.5)

pm2

pm3 <- DimPlot(Macrophage, reduction = "scvi", group.by = "seurat\_clusters", label = F, pt.size = 1.5)

pm3

pm4 <- FeaturePlot(Macrophage, features = c("FCN1","FABP4","MARCO", "INHBA","FOLR2","LYVE1","LGMN","CD163",

 "TYMS","MKI67", "TOP2A", "NUSAP1"),

 min.cutoff = "q10", reduction = "scvi")

pm4

pm5 <- VlnPlot(Macrophage, features = c("FCN1","FABP4","MARCO", "INHBA","FOLR2","LYVE1","LGMN","CD163",

 "TYMS","MKI67", "TOP2A", "NUSAP1"),

 group.by = "RNA\_snn\_res.0.6",

 stack = TRUE, flip = TRUE)

pm5

Macrophage\_markers <- FindAllMarkers(Macrophage, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25, reduction = "scvi")

write.csv(Macrophage\_markers, file = "Macrophage Markers res0.6.csv")

clusters\_to\_merge <- c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)

Macrophage$custom\_cluster <- Idents(Macrophage)

new\_cluster\_label\_1 <- "Monocyte"

new\_cluster\_label\_2 <- "AM"

new\_cluster\_label\_3 <- "IM"

new\_cluster\_label\_4 <- "Prolif.M"

Macrophage$custom\_cluster <- factor(

 Macrophage$custom\_cluster,

 levels = c(levels(Macrophage$custom\_cluster), new\_cluster\_label\_1, new\_cluster\_label\_2,

 new\_cluster\_label\_3, new\_cluster\_label\_4))

Macrophage$custom\_cluster[Macrophage$custom\_cluster %in% c(4, 7, 12)] <- new\_cluster\_label\_1

Macrophage$custom\_cluster[Macrophage$custom\_cluster %in% c(0, 2, 5, 9, 11, 13)] <- new\_cluster\_label\_2

Macrophage$custom\_cluster[Macrophage$custom\_cluster %in% c(1, 3, 6, 8, 10, 14)] <- new\_cluster\_label\_3

Macrophage$custom\_cluster[Macrophage$custom\_cluster %in% c(15)] <- new\_cluster\_label\_4

pm6 <-DimPlot(Macrophage, group.by = "custom\_cluster", reduction = "scvi")

pm6

pm7 <-DotPlot(Macrophage, features = c("MAFB","IL10", "LGMN","CD163","FOLR2","LYVE1","SPP1","CCL18","CCL2",

 "MS4A6A","CXCL8","CYP1B1","IL4I1","LDLR","CD300E","BATF"),

 group.by = "custom\_cluster", dot.min = 0, dot.scale = 25,cols = c("blue","red")) + RotatedAxis()

pm7

pm8 <-DotPlot(Macrophage, features = c("FABP4","INHBA","GSN","FBP1",

 "PPARG","CXCL5","AXL","AQP3","CLEC4E",

 "ABCG1","CITED2"),

 group.by = "custom\_cluster", dot.min = 0, dot.scale = 25, cols = c("blue","red")) + RotatedAxis()

pm8

bulk <- AggregateExpression(Macrophage, group.by = c("custom\_cluster","orig.ident"),return.seurat = F)

bulk <- bulk$RNA

bulk1 <- as.data.frame(bulk)

write.csv(bulk1, file = "pseudoRNAseq Macrophages.csv")