cTPnet surface protein abundance inference

2022-09-12

load the libraries and the python env for cTP-net

```
library(cTPnet)
library(Seurat)
library(reticulate)
#conda env for cTPnet
use_condaenv("~/anaconda3/envs/cTPnet/")
```

load the scRNA-seq dataset the pre-trained model

Load the scRNA-seq dataset for an AML patient, which is enriched for progenitor and later stage cells. You can find more for this dataset in Fig 4 of the cTP-net paper https://doi.org/10.1038/s41467-020-14391-0.

```
# load scRNA-seq data
amlData<-read.table("process/GSM3587990_AML921A-D0.dem.txt",
                    sep = "\t",stringsAsFactors = F,row.names = 1,header = T)
# build Seurat object
amlObj = CreateSeuratObject(amlData)
# load the pre-trained model for 24 surface proteins
amlObj = cTPnet(amlObj,data_type="Seurat3",model_file_path="support/cTPnet_weight_24")
```

```
## Start data preprocessing...
## Start imputation. Running python ...
## Postprocess...
## Done!
```

data normalization and feature selection

```
# normalization
amlObj <- NormalizeData(amlObj, display.progress = FALSE)
# choose variable features
amlObj <- FindVariableFeatures(amlObj, do.plot = FALSE)
# scaling the dataset
amlObj <- ScaleData(amlObj, display.progress = FALSE)
```

dimention reduction and clustering

```
# PCA
amlObj <- RunPCA(amlObj, verbose = FALSE)
# find neighboring cells
amlObj <- FindNeighbors(amlObj, dims = 1:25, k.param = 20)
# define clusters
amlObj <- FindClusters(amlObj, resolution = 0.8)
```

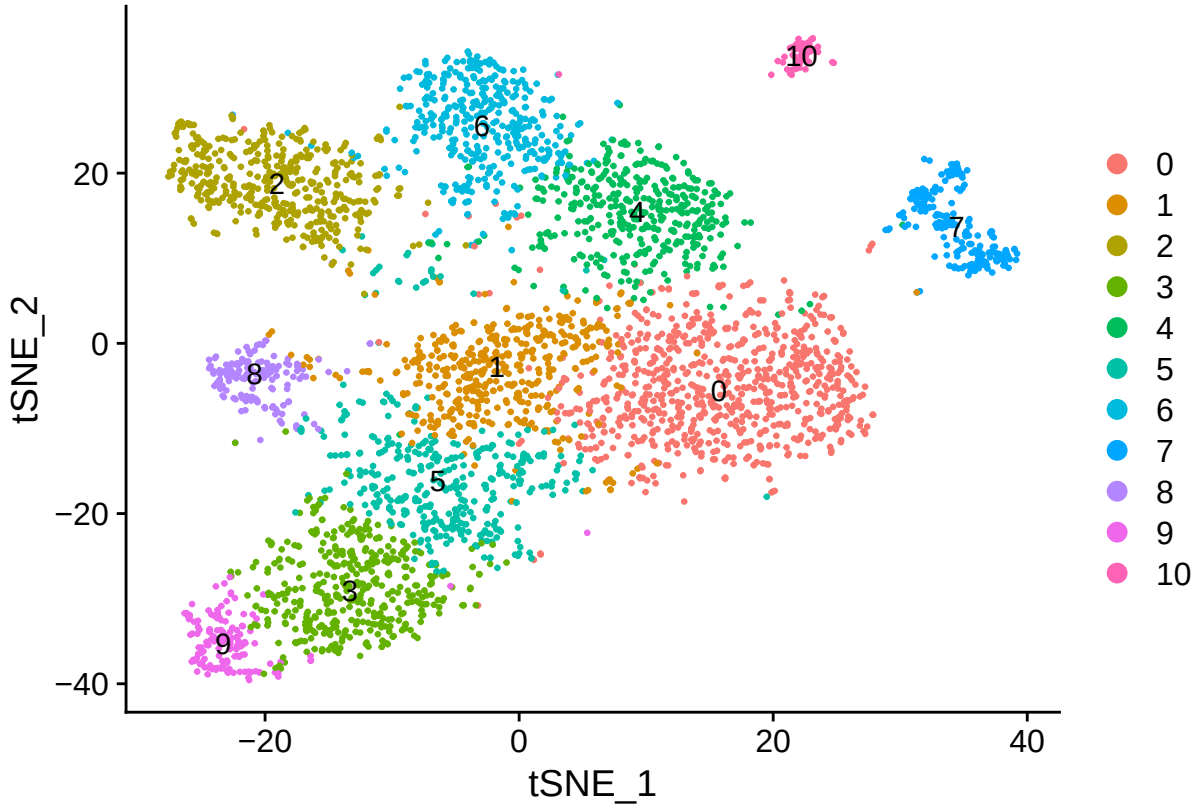
```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 3813
## Number of edges: 158681
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.8258
## Number of communities: 11
## Elapsed time: 0 seconds
```

```
# t-SNE dimension reduction
amlObj <- RunTSNE(amlObj, dims = 1:25, method = "FIt-SNE", max_iter=2000)
```

cluster and marker genes visualization

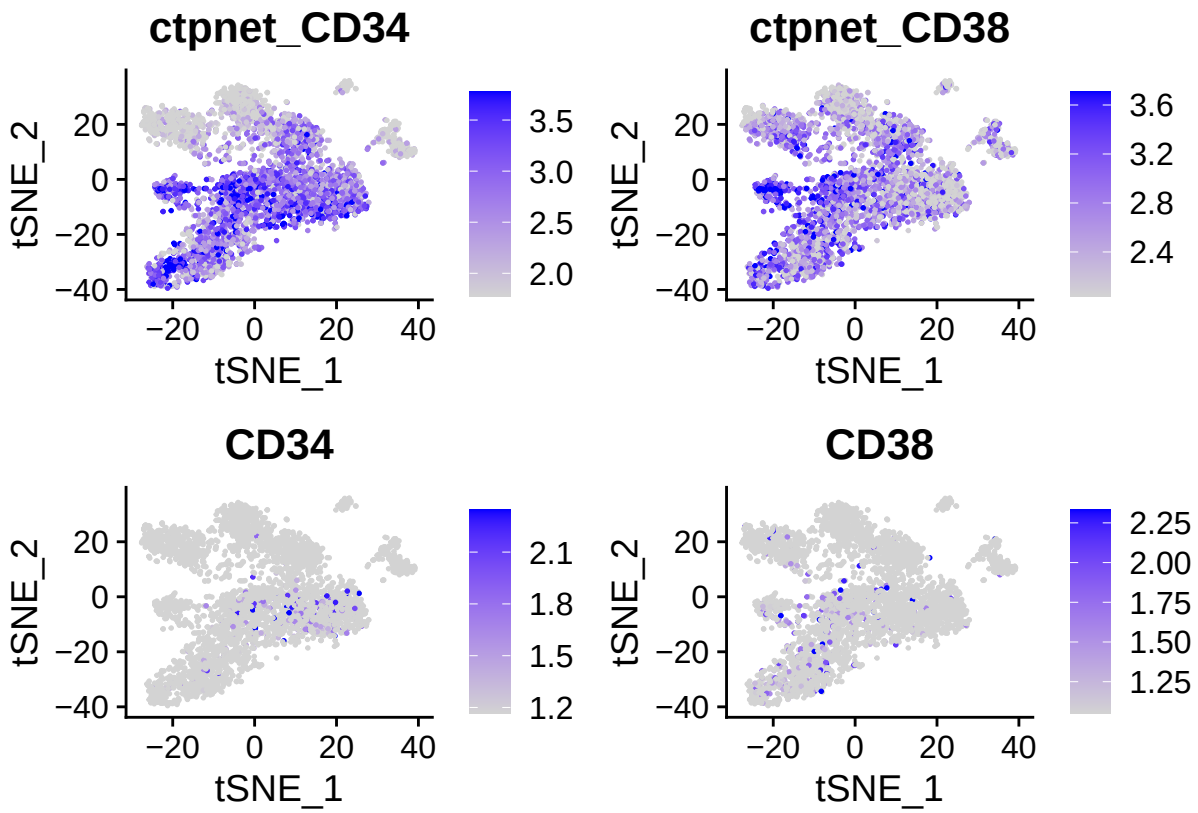
Plot the clusters and two key marker genes' abundance in both mRNA and inferred protein level for progenitor cells and blast cells. These are the surface markers for sorting the cells in the single cell MS paper https://doi.org/10.1038/s41467-021-23667-y.

```
# clusters
DimPlot(amlObj, label = TRUE, pt.size = 0.5)
```




```
# marker genes indicating progenitor and blast cells
FeaturePlot(amlObj, features = c(
```

```
  "ctpnet_CD34", "ctpnet_CD38",
  "CD34", "CD38"
), min.cutoff = "q25", max.cutoff = "q95", ncol = 2, pt.size=0.3)
```



select progenitor and blast cells enriched clusters

As shown in the plots, the inferred protein abundance are more informative for defining cell types. Here, cluster 6 and cluster 2 cells are selected for representing progenitor and blast cells, respectively.

```
# cluster 6 for progenitor cells
aml.prog.barcode<-rownames(subset(amlObj@meta.data,seurat_clusters==6))
# cluster 2 for blast cells
aml.blast.barcode<-rownames(subset(amlObj@meta.data,seurat_clusters==2))
```

extract average RNA/protein abundance

The average RNA/protein abundance of all 24 surface proteins are extracted from the two cell types.

```
# progenitor cells
prt.prog.marker<-rowSums(amlObj@assays$cTPnet@data[,aml.prog.barcode])/length(aml.prog.barcode)
rna.prog.marker<-rowSums(amlObj@assays$RNA@data[,aml.prog.barcode])/length(aml.prog.barcode)

aml.marker.names<-names(prt.prog.marker)
rna.prog.marker<-rna.prog.marker[aml.marker.names]
```

```r
aml.marker<-rbind(rna.prog.marker,prt.prog.marker)
write.table(aml.marker,file = "process/progCell.marker.txt",sep = ",",quote = F,row.names = T)

# blast cells
prt.blast.marker<-rowSums(amlObj@assays$cTPnet@data[,aml.blast.barcode])/length(aml.blast.barcode)
rna.blast.marker<-rowSums(amlObj@assays$RNA@data[,aml.blast.barcode])/length(aml.blast.barcode)

rna.blast.marker<-rna.blast.marker[aml.marker.names]
aml.marker<-rbind(rna.blast.marker,prt.blast.marker)
write.table(aml.marker,file = "process/blastCell.marker.txt",sep = ",",quote = F,row.names = T)
```