Doublet finder

Ricardo Martins-Ferreira

2023-07-06

Example for the Grubman et al. 2019 (PMID: 31768052)

```
##        Seurat DoubletFinder
##        FALSE        FALSE
```

The pre-processed Seurat object (including all cell types) previously obtained for each dataset/study (described at "Individual dataset processing.pdf") is splited into individual Seurat objects for each of the individual subjects/samples.

Doublet finder is applied to each subject/sample individually.

Doublet finder loop (by Subject)

```r
# list subjects
mylist <- unique(Seurat@meta.data$Subject)

for (i in 1:length(mylist)) {

Seurat_sample <- Seurat[,Seurat@meta.data$Subject==mylist[i]]

# run sctransform
Seurat_sample <- SCTransform(Seurat_sample, verbose = FALSE,
                                    conserve.memory = FALSE)

# Perform linear dimensiona reduction
Seurat_sample <- RunPCA(Seurat_sample, verbose = FALSE)

#Dimensionality reduction and clustering
Seurat_sample <- RunUMAP(Seurat_sample,  reduction = "pca",dims = 1:30, verbose = FALSE)
Seurat_sample <- FindNeighbors(Seurat_sample,  reduction = "pca",dims = 1:30)
Seurat_sample <- FindClusters(Seurat_sample,resolution = 0.02)

## pK Identification (no ground-truth)
sweep.res.list_Seurat <- paramSweep_v3(Seurat_sample, PCs = 1:10, sct = T)
sweep.stats_Seurat <- summarizeSweep(sweep.res.list_Seurat, GT = FALSE)
bcmvn_Seurat <- find.pK(sweep.stats_Seurat)
df <- bcmvn_Seurat[which.max(bcmvn_Seurat$BCmetric),]

max <- as.character(df[,"pK"])
```

```r
## Homotypic Doublet Proportion Estimate
annotations <- Seurat_sample@meta.data$SCT_snn_res.0.02

homotypic.prop <- modelHomotypic(annotations)
## ex: annotations <- seu_Seurat@meta.data$ClusteringResults

nExp_poi <- round(0.075*nrow(Seurat_sample@meta.data))
## Assuming 7.5% doublet formation rate - tailor for your dataset

nExp_poi.adj <- round(nExp_poi*(1-homotypic.prop))

## Run DoubletFinder with varying classification stringencies
Seurat_sample <- doubletFinder_v3(Seurat_sample, PCs = 1:10, pN = 0.25,
                                  pK = as.numeric(max), nExp = nExp_poi,
                                  reuse.pANN = FALSE, sct = T)
colnames(Seurat_sample@meta.data)[ncol(Seurat_sample@meta.data)] <- "DoubletFinder"

doublets<- Seurat_sample@meta.data[Seurat_sample@meta.data$DoubletFinder=="Doublet",]
  write.csv(doublets %>% rownames(), paste0("Doublets_",mylist[i],".csv"), row.names = F)
}
```