

- 2024/05/13-2024/0519
 - 2024/05/16
 - 整理基于R语言的单细胞数据分析--preprocessing部分
 - 1.实验目的
 - 2.实验内容
 - 3.实验材料：
 - 3.1数据来源：
 - 3.2软件与平台：
 - 4.实验步骤
 - 4.1Load raw data
 - 4.2Quality control
 - 4.3Standard work flow
 - 5.注意事项
 - 2024/05/17
 - 整理基于Python的单细胞数据分析--Seurat.object与Scanpy.anndata格式转换
 - 1.实验目的
 - 2.实验内容
 - 3.实验材料
 - 3.1数据来源：
 - 3.2软件与平台
 - 4.实验步骤
 - 4.1SCP
 - 4.2zellkonverter
 - 4.3reticulate
 - 4.4SeuratDisk
 - 4.5使用Scanpy.read_10x_mtx()直接读取原始glioma数据
 - 5.实验结果

2024/05/13-2024/0519

2024/05/16

整理基于R语言的单细胞数据分析-- preprocessing部分

1.实验目的

1.1熟悉基于R的seurat标准预处理分析流程

1.2理解seurat object数据结构

1.3为后续细胞注释分析工作作准备

2.实验内容

基于数据库下载的单细胞数据，使用R包Seurat进行标准预处理流程，跑通并整理成可读性强的脚本。

3.实验材料：

3.1数据来源：

Single-cell analysis of human glioma and immune cells identifies S100A4 as an immunotherapy target(GSE182109)

Single Cell Portal数据，人，脑胶质瘤GBM, T cells (subset)

链接: https://singlecell.broadinstitute.org/single_cell/study/SCP1985/single-cell-analysis-of-human-glioma-and-immune-cells-identifies-s100a4-as-an-immunotherapy-target-gse182109

3.2软件与平台：

R (v4.3.3) ; RStudio; Seurat (v5.0.3, <https://github.com/satijalab/seurat>) ; Seurat tutorial (<https://satijalab.org/seurat/>)

4.实验步骤

简单介绍预处理流程

4.1Load raw data

```
1.read10X / readhd5 / readRDS

2.subset my interested cell type: T cells
```

4.2Quality control

```
1.add percent.mt

2.subset according to percent.mt , nFeature_RNA and nCount_RNA
```

4.3Standard work flow

```
1.Normalization

2.FindVariableFeatures

3.Scale

4.PCA

5.FindNeighbours

6.FindClusters (we can setup resolution e.g. res=0.1, 0.8)

7.UMAP

8.Finally, we will get a clustering map

Remember: if the data comes from muti datasets, do integration after PCA(e.g. CCA-Integration, RPCA-Integration)
```

5.注意事项

- 1.抽取数据时，尽量少用subset函数，此函数不能多次进行子集操作
- 2.做质量控制时，务必设定好筛选条件：nFeature_RNA, nCount_RNA, Percent.mt。可通过绘制Vlnplot可视化直观的了解各自的分布。
- 3.标准流程不必多说，再做之前，要了解数据来源，是否是多个数据集，考虑批次差异的影响，应进行Integrate Layers操作，而后方可进行PCA等后续降维聚类。
- 4.对于降维聚类，即FindClusters和UMAP，需要设定不同的分辨率 (resolution)，因为后续的注释是一个极其费时费力的工作，需要先以低倍看整体大群（例如 res=0.1），而后以高倍看细分小群（例如res=0.8）
- 5.运行过程中，需要保持注释的好习惯，尽量用英文注释，并注意划分功能段落，增加可读性的同时也能

锻炼自己的英语水平

6.对于过程中产生的临时文件，对于需要的数据，例如大文件（分析时间过长），中部关键数据，要随时保存，并写下读取代码。

7.变量的命名，在保证可读的基本要求下，尽量简洁。（例如：读取的原始单细胞counts矩阵可命名为：sc.raw.counts）

2024/05/17

整理基于Python的单细胞数据分析-- Seurat.object与Scanpy.anndata格式转换

1.实验目的

1.1了解Seurat.object与Scanpy.anndata格式转换的方法

1.2尝试读取2024/05/16预处理的SingleCellPortal单细胞数据（即sc.tcells）

2.实验内容

使用R语言，读取seurat标准流程预处理的单细胞数据（即sc.tcells.after_umap.rpca~without QC.rds），并将其转换为Scanpy.anndata格式，以便后续用python处理单细胞数据

3.实验材料

3.1数据来源：

2024/05/16基于R包Seurat预处理的Tcells 单细胞数据，该数据已经过RPCA整合，并进行了UMAP降维聚类

3.2软件与平台

R (v4.3.3) ; RStudio; R包Seurat (v5.0.3, <https://github.com/satijalab/seurat>) ;
Python库Scanpy (v1.10.1, <https://github.com/scverse/scanpy>) ; Scanpy tutorial
(<https://scanpy.readthedocs.io/en/stable/#>) ; AnnData tutorial

4.实验步骤

4.1SCP

```
library(SCP)

data("pancreas1k")

adata <- srt_to_adata(pancreas1k)

adata$write_h5ad("pancreas1k.h5ad")
```

4.2zellkonverter

```
sce_obj <- as.SingleCellExperiment(sce_obj, assay = c("RNA"))

library(zellkonverter)

writeH5AD(sce_obj, "sce_obj.h5ad", X_name = 'counts')
```

4.3reticulate

```
require(Seurat) require(reticulate)

seu <- readRDS('your_path_seurat_object_rds')

#load python anndata package

anndata <- reticulate::import('anndata')

#create anndata object

adata <- anndata$AnnData(X = seu@assays$RNA@layers$counts, obs =
data.frame(row.names = rownames(seu)), var = seu@meta.data )
adata$write("your_path_scanpy_obj_h5ad")

#Of note that adata require an inversion in python scanpy

import scanpy as sc

adata = sc.read_h5ad("your_path_scanpy_obj_h5ad")
```

```
adata = adata.T
```

4.4 SeuratDisk

```
library(SeuratDisk)
```

```
SaveH5Seurat(sc.tcells, filename = "./tmp/pbmc3k.h5Seurat")
```

```
Convert("./tmp/pbmc3k.h5Seurat", dest = "h5ad")
```

4.5 使用 Scanpy.read_10x_mtx() 直接读取原始 glioma 数据

```
adata=sc.read_10x_mtx("./data/raw/")
```

5. 实验结果

1. 上述四种转换方法尝试均失败

2. 使用 Scanpy 直接读取 raw.data.glioma 也会出现报错