

Step 系列：scRNA-seq 基本分析

2024-01-01

LiChuang Huang



@ 立效研究院

Contents

1 Step 系列共性特征	1
1.1 理念	1
1.2 泛用方法	2
1.2.1 提取方法	2
1.2.2 存储方法	2
1.2.3 空间管理	2
1.2.4 查看 step 方法的默认参数	2
1.2.5 额外的信息	3
1.3 关于安装配置	3
1.4 关于本文档	4
2 Step 系列: scRNA-seq 基本分析	4
2.1 摘要	4
2.1.1 目的	4
2.1.2 解决的问题	4
2.2 方法	4
2.3 安装 (首次使用)	4
2.3.1 安装依赖	4
2.3.1.1 安装 Seurat v5	4
2.3.1.2 安装 seurat-wrappers	4
2.3.1.3 安装 monocle3	4
2.3.1.4 安装 CellChat	5
2.3.1.5 其它程序	5
2.3.2 安装主体	5
2.4 使用说明	5
2.5 示例分析	5
2.5.1 数据准备	5
2.5.2 分析流程	5
2.5.2.1 Job-seurat	5
2.5.2.2 Step1	5
2.5.2.3 Step2	5
2.6 技巧	5

List of Figures

List of Tables

1 Step 系列共性特征

1.1 理念

生物信息分析工具种类繁多，开发者出于各种原因，导致工具的适用性千差万别。学习和应用这些工具的成本可能是高昂的（甚至它们可能有一些不为人知的漏洞），想要把它们串连在一起分析，更是要付出一些分析之外的代价：像开发者一样调试这些程序。

Step 系列的分析方法为每一个制定好的方法、思路或工具（一般是领域中的权威、经典或翘楚）设立统一使用的标准，统一的数据存储，统一的应用流程，大幅度降低了学习成本、使用成本、应用成本。通过避免大量“分析之外”的繁琐工作，达到提高分析效率的目的。

Step 系列分析方法的一些基本特性：

- 统一的分析平台。Step 背后涵盖的工具可能涉及：R 包、Python 包、Java 包、Linux 命令行工具等。但最终用于分析的始终是 R 语言。Step 系列的方法通过在 R 中调用各种分析工具，实现不同分析平台之间的工具调用和数据对接（降低了学习和应用成本）。
- 统一的数据存储单位。Step 系列所有的分析流程（workflow）都以一个对象（Object）存储。这避免了如果一个流程中涉及纷繁复杂的分析方法，人工存储数据（中间数据、最终输出的图片、表格）出错的可能（提高效率）。必要时，通过统一的方法提取这些数据（就像在图书馆的某一层的某一排的书架的某一个柜子取出一本书）。
- 统一的方法名称。Step 系列的所有 Workflow 的方法名称都是统一的（step1、step2、step3..... map、vis 等），但不会因为一同调用而出错。这是为了减轻分析者的负担而设计的（如果一次完整分析涉及十几种工具，每个工具又有十几种方法名称，那对分析者的记忆量和细心度的考验是惊人的）（提高效率）。
- 规范的分析流程。大多数的分析工具本身具备多种方法以适用灵活分析，但也提高了学习、应用成本。Step 系列的各个 Workflow 的流程都是单向的，对分析方法或思路的组合应用大都是建立在官方指南或教程的基础上，又或者是泛用性（降低了学习成本）。如果有不同思路，可以通过关键参数调整方法，或者视情况搭建一个额外的 Workflow。因此，每个 Workflow 是以分析思路为分门别类的，而不是工具本身。
- 提供权威、经典、泛用的分析方法。Workflow 创建前，会广泛查阅时下文献，对工具择优而选、择新而用（比如，更趋向于选择更新的来自于 Nature Biotechnology 的分析方法）。
- 提供泛用的组合思路。通过组合各种数据库、分析工具，应对千奇百怪的分析需求。同时，一些适当的组合，会发挥超出单一工具的价值（Fig. ?? 的右图提供了现如今存在的许多组合思路；同时，每个 Workflow 内部又存在一些思路组合）。
- 不断开发进化。每一个完成的 Workflow 不是固定不变的，在面临新的分析环境、新的分析需求，更多的功能会被加入到“step”方法中，让每一次的进步都直接应用于今后所有的同类分析。另外，一些大大小小的创新（比如，更严谨或更酷炫的绘图）也会不断追加到方法中。
- 效率至上。所有分析方法以输入命令形式实现，允许批量处理。
- 附带实用工具。分析流程相对固定，但通过提供一些工具（比如用于额外的绘图），可以将分析更加灵活。

1.2 泛用方法

Input

```
sr <- job_seurat("./data")
sr <- step1(sr)
sr <- step2(sr)
sr <- step3(sr)
sr <- step4(sr)
sr <- step5(sr)
sr <- step6(sr)

mn <- asjob_monocle(sr)
mn <- step1(mn)
mn <- step2(mn)
mn <- step3(mn)
```

1.2.1 提取方法

1.2.2 存储方法

1.2.3 空间管理

1.2.4 查看 step 方法的默认参数

step 方法的参数力求精简，一般只保留关键的参数用以控制分析。但这些参数可能会在将来被拓展 (添加额外的参数) 以适应新的分析需求。

可以通过类似以下方式查看默认参数：

Input

```
## 示例: seurat 工作流的 step1 的默认参数
not(.job_seurat())
step1
```

```
## job_seurat:
```

```
##      x
```

```
##
```

```
## -- Methods parameters -----
```

Input

```
## 示例: monocle 工作流 step2 的默认参数
not(.job_monocle())
step1
```

```
## job_monocle:
```

```
##      x, groups = x@params$group.by, pt.size = 1.5, pre = F, norm_method= "none"
```

```
##
```

```
## -- Methods parameters -----
```

1.2.5 额外的信息

工作流创建时参考的信息源参考文献信息

Input

```
# 创建一个空的 Seurat 工作流对象 'sr_1'
sr_1 <- .job_seurat()
# 查看方法说明
sr_1@method
```

```
## [1] "The R package `Seurat` used for scRNA-seq processing; `SCSA` (python) used for cell type annota
```

Input

```
# 查看官方网站或信息源网站
sr_1@info
```

```
## [1] "Tutorial: https://satijalab.org/seurat/articles/pbmc3k\_tutorial.html"
```

1.3 关于安装配置

所有 Step 系列的方法的安装配置会尽可能详细的罗列，但由于笔者（开发者）仅在自身的 Linux (Ubuntu 发行版) 系统做过调试，并不确定在其它的机器上会遇到哪些安装的特殊问题。如有疑问，或安装上的困难，请联系：Huang Lichuang (huanglichuang@wie-biotech.com)。

1.4 关于本文档

2 Step 系列: scRNA-seq 基本分析

2.1 摘要

2.1.1 目的

2.1.2 解决的问题

2.2 方法

Mainly used method:

- Other R packages (eg., `dplyr` and `ggplot2`) used for statistic analysis or data visualization.

2.3 安装 (首次使用)

2.3.1 安装依赖

2.3.1.1 安装 Seurat v5

Input

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
if (!requireNamespace("remotes", quietly = TRUE))
  install.packages("remotes")
install.packages(c("sva"))
BiocManager::install(c('SparseArray', 'fastDummies', 'RcppHNSW', 'RSpectra'))
remotes::install_github("satijalab/seurat", "seurat5")
```

2.3.1.2 安装 seurat-wrappers

Input

```
remotes::install_github('satijalab/seurat-wrappers')
```

2.3.1.3 安装 monocle3

Input

```
BiocManager::install(c('BiocGenerics', 'DelayedArray', 'DelayedMatrixStats',
  'limma', 'lme4', 'S4Vectors', 'SingleCellExperiment',
  'SummarizedExperiment', 'batchelor', 'HDF5Array',
  'terra', 'ggrastr', 'rsample'))
remotes::install_github('cole-trapnell-lab/monocle3')
```

2.3.1.4 安装 CellChat

Input

```
BiocManager::install(c('NMF', 'circlize', 'ComplexHeatmap', 'BiocNeighbors'))
remotes::install_github("sqjin/CellChat")
```

2.3.1.5 其它程序

以下可能是其它需要安装的程序：

Input

```
install.packages("reticulate")
reticulate::install_miniconda()
reticulate::py_install(packages = 'umap-learn')
```

2.3.2 安装主体

2.4 使用说明

2.5 示例分析

2.5.1 数据准备

2.5.2 分析流程

2.5.2.1 Job-seurat

2.5.2.2 Step1

参数说明运行结果参考函数

2.5.2.3 Step2

测试

2.6 技巧