

单细胞 SingleR 结果说明文档

1. 背景介绍

单细胞亚群的鉴定一直以来是单细胞转录组分析的一个难点但是有非常重要的部分，目前单细胞亚群自动化鉴定的工具种类繁多，其中 SingleR 是目前单细胞亚群鉴定中最常用的一种，SingleR 是 Aran 等人与 2019 年在 Nature Immunology 发表的一篇关于专门用于免疫微环境中的工具，其大概思路就是将单细胞表达矩阵与纯的 bulk RNA 表达数据进行相似性分析，然后根据相似性最大的进行细胞类型判断，支持人和小鼠两个物种。

SingleR 开始的版本对外来输入数据支持度不太友好，而且运行比较慢，但是自从升级 1.0.0 以上版本一行，其运行方式与前面的方式完全不一样，对外来的数据集支持度大大的提高，并且官网支持的有 7 个数据集，数据集的细胞类型分的更细，能够到 CD4 Naïve 细胞类型，老师可以根据研究目的的不同，选择不同的数据集，数据集有专门的免疫数据集或者其他数据集，具体见下表格：

Retrieval function	Organism	Samples	Sample types	No. of main labels	No. of fine labels	Cell type focus
HumanPrimaryCellAtlasData()	human	713	microarrays of sorted cell populations	37	157	Non-specific
BlueprintEncodeData()	human	259	RNA-seq	24	43	Non-specific
DatabaseImmuneCellExpressionData()	human	1561	RNA-seq	5	15	Immune
NovershternHematopoieticData()	human	211	microarrays of sorted cell populations	17	38	Hematopoietic & Immune
MonacoImmuneData()	human	114	RNA-seq	11	29	Immune
ImmGenData()	mouse	830	microarrays of sorted cell populations	20	253	Hematopoietic & Immune
MouseRNAseqData()	mouse	358	RNA-seq	18	28	Non-specific

上面的一共 7 个数据集，其中人有 5 个，小鼠两个，人和小鼠都有免疫数据集，因此这个工具对免疫数据支持度较好，其他的数据集根据研究目的不同，有不同的特点。具体的数据说明可见官网介绍：http://www.bioconductor.org/packages/release/bioc/vignettes/SingleR/inst/doc/SingleR.html#5_available_references。

2. 结果说明

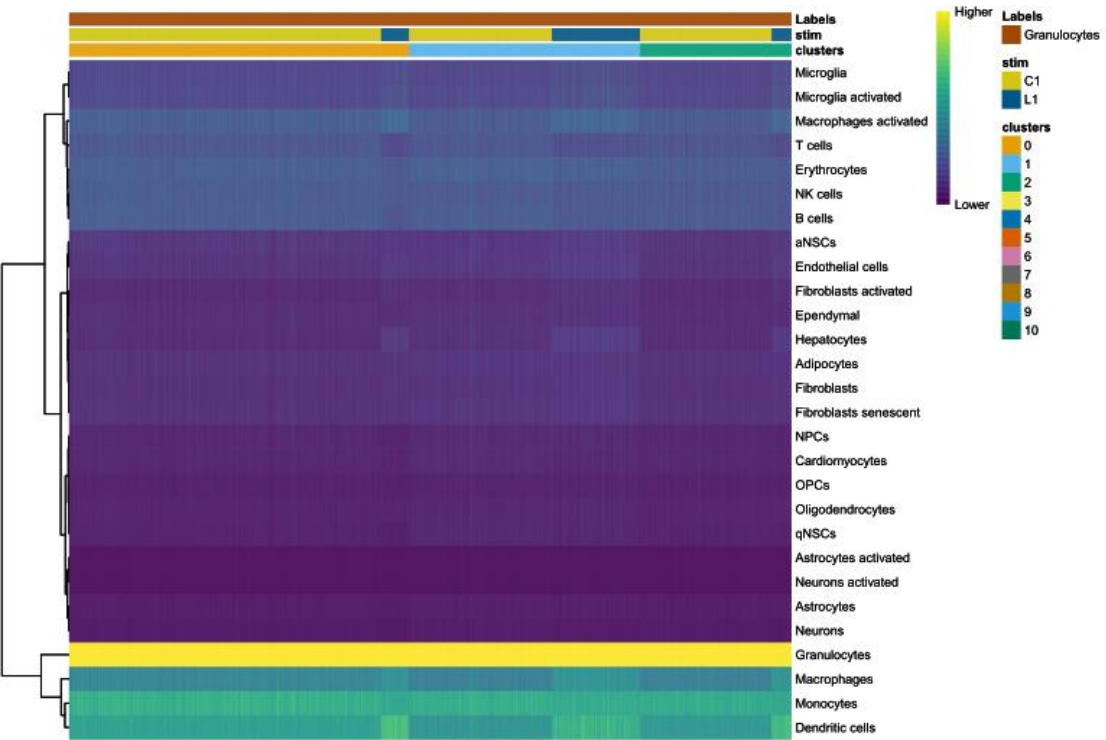
目前的结果主要包含三个文件：

```
result/  
├── 前缀_singleR_celltype.xls  
├── 前缀_singleR.rds  
├── 前缀_stim_celltype.pdf  
├── singleR_celltype_stat.xls  
└── singleR_celltype_stat_top1.xls
```

前缀_singleR_celltype.xls，这个文件为每个细胞的细胞类型预测结果文件，每一行为一个细胞，每一列为一个细胞类型的得分，示范结果如下图所示，这里可以看到每个细胞的类型。其实这个结果也可以放在 seurat 里面，然后画一个细胞类型的散点图可以可以。

Barcode	scores.Adipocytes	scores.aNSCs	scores.Astrocytes
AAACCTGAGCGTGTCC. C1	0.131260178	0.143637183	0.054680223
AAACCTGAGGTGGGTT. C1	0.136119448	0.142036531	0.060962082
AAACGGGAGCGCCTCA. C1	0.12290603	0.132774499	0.045693705
AAACGGGAGGCTAGAC. C1	0.143540786	0.161956683	0.074074893
AAACGGGTTAGTGAAT. C1	0.123233247	0.145603098	0.058439447
AAACGGGTCCAAAGTC. C1	0.119071358	0.117030362	0.040299838
AAAGATGAGGATATAC. C1	0.128693195	0.133108561	0.048321615
.....

前缀_stim_celltype.pdf, 这个结果为其主要的结果，主要是用热图展示了其每个细胞、每个亚群、每个样品的细胞类型的可能性结果图，黄色代表可能性越大，深蓝代表可能性越低，热图的上方的三个标签，clusters 代表亚群，stim 代表不同的样品，labels 代表软件自动鉴定的细胞类型，这里的是挑选了一部分数据进行自动化鉴定的，因此其结果大部分为**粒细胞**。



前缀_singleR.rds, 这个文件主要 SingleR 的结果文件，这是一个二进制文件，这个一般需要 R 才能打开 (result<-readRDS(‘前缀_singleR.rds’)), 如果老师对做的过程感兴趣，可以打开这个文件，然后通过 R 进行展示一些相关信息，如果老师对 R 不熟悉，不建议打开此文件。

singleR_celltype_stat.xls 文件为依据 singleR 的预测结果的统计的总表

- 1) clusters: 类群
- 2) labels: 细胞类型
- 3) Barcode: 预测属于该类型的细胞数量
- 4) cluster_totalcells: 该 cluster 的总的细胞数
- 5) precent: 该细胞类型占该类群总数的比例

pbmc_singleR_celltype_stat_top1.xls 为上表中占比最多的细胞类群定义为注释类型, 列名同上.

3. 参考文献

[Aran, Looney, Liu et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature Immunology \(2019\)](#)

官方说明文档: [http://www.bioconductor.org/packages/release/bioc/vignettes/SingleR/inst/doc/SingleR.html#5 available references](http://www.bioconductor.org/packages/release/bioc/vignettes/SingleR/inst/doc/SingleR.html#5_available_references)