velocyto:细胞 RNA表达平衡变化分析

项目名称: velocyto—细胞 RNA表达平衡变化分析

项目负责:北京寻因生物科技有限公司

撰写人:谢卓明，信息分析工程师

2021 年 07 月 30 日

特别声明

本文档属于内部资料，仅提供给公司正式员工参阅;未经允许限制资料的传输，禁止公开传阅。

版本历史记录

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 日期 | 提交人 | 备注 |
| V1 | 2021.7.30 | 谢卓明 | velocyto:细胞 RNA表达平衡变化分析 |
|  |  |  |  |

版权所有:北京寻因生物科技有限公司保留所有权利。

本产品或文档受版权保护，其使用、复制、发行和反编译均受许可证限制。未经北京寻因生物科技有限公司及其授权者的事先书面许可，不得以任何形式、任何手段复制本产品或文档的任何部分。

**目录**

[velocyto:细胞 RNA表达平衡变化分析 1](#_Toc78557546)

[项目名称: velocyto—细胞 RNA表达平衡变化分析 1](#_Toc78557547)

[项目负责:北京寻因生物科技有限公司 1](#_Toc78557548)

[撰写人:谢卓明，信息分析工程师 1](#_Toc78557549)

[2021 年 07 月 30 日 1](#_Toc78557550)

[1 文档说明 4](#_Toc78557551)

[1.1 关于手册 4](#_Toc78557552)

[1.2 面向的使用者 4](#_Toc78557553)

[1.3 信息反馈 4](#_Toc78557554)

[2 背景介绍 5](#_Toc78557555)

[3 流程使用说明 6](#_Toc78557556)

[3.1 模块简介 6](#_Toc78557557)

[3.2 模块参数描述 6](#_Toc78557558)

[3.2.1 配置文件 8](#_Toc78557559)

[3.2.2 脚本配置 8](#_Toc78557560)

[3.3 执行过程 9](#_Toc78557561)

[4 结果展示及解读 10](#_Toc78557562)

[5 文献实例 11](#_Toc78557563)

[6 参考文献 12](#_Toc78557564)

[7 问题反馈及建议 12](#_Toc78557565)

# 文档说明

## 关于手册

该文档旨在基于含有序列标签单细胞转录(sc-RNA)数据，利用基因表达矩阵进行机器学习分析，进而模拟时间发育过程的动态变化。

## 面向的使用者

寻因信息分析工程师。

## 信息反馈

如果对本文档有任何意见和建议，请联系:xiezhuoming@seekgene.com

# 背景介绍

单细胞 RNA 测序可以以高定量准确度、灵敏度和通量来揭示 RNA 丰度。然而，这种方法仅捕获某个时间点的静态快照，无法在时间上重现（如胚胎发生或组织再生）发育过程的RNA表达平衡变化。Velocyto 可以通过在常见的单细胞 RNA 测序方案中区分未剪接和剪接的 mRNA 来直接估计 RNA 表达平衡变化。

# 流程使用说明

## 模块简介

脚本名称: velocyto.prep.R、velocyto.plot.R

脚本版本:v1.0

脚本路径: /PROJ/development/Seek\_Script/velocyto

## 模块参数描述

velocyto.prep.R

|  |  |
| --- | --- |
| 参数 | 描述 |
| --bam | STAR 比对后的 bam 格式文件。可在标准流程结果路径，eg：alignmentYaml/SAMPLE/STAR/SAMPLE\_SortedByCoordinate.bam（seekone）；cellrangerYaml/SAMPLE/outs/possition\_sorted.bam(10x)找到。 |
| --filtered\_barcode | 初步过滤后的 barcodes 文件。可在在标准流程结果路径下，eg：countBarcodeYaml/SAMPLE/filtered\_feature\_bc\_matrix/barcodes.tsv.gz（seekone）；cellrangerYaml/SAMPLE/outs/filtered\_feature\_bc\_matrix/barcodes.tsv.gz(10x)找到。 |
| --species | 目前支持 human|mouse|rat，可在./ref文件夹中自定义 gtf 文件，命名为 SPECIES\_genes.gtf。 |
| --add\_tag | T|F，目前的 seekone 标准流程比对结果 bam 文件没有 CB\UB tag，选择 T 时会加上，10x 则不需要，使用没有相关 tag 的 bam 文件会报错。 |
| --outdir | 默认是 ./output |

velocyto.plot.R

|  |  |
| --- | --- |
| 参数 | 描述 |
| --velocyto\_rds | 包含 4 个 assay 的 seurat 对象（velocyto.rds）。可在 velocyto.prep.R 的结果文件中找到。 |
| --seurat\_obj2project | 用于画图的 seurat 对象，rds格式。eg：basicSeuratYaml/SAMPLE/SAMPLE\_seurat.rds。 |
| --outdir | 输出文件路径，默认是./output。 |
| --reduction\_space | seurat\_obj 的降维空间名称，默认是 umap。 |
| --default\_assay | seurat\_obj 的 default\_assay 名称，默认是 RNA。 |
| --gene2show | 需要画图的基因列表，多个基因可以,分隔，没有 fit 出 γ 值的基因不会出图。 |
| --old\_fit | oldfit.rds。不再处理原始 velocyto.rds ，使用已有拟合结果，可用于降低画图时间。 |
| --n.cores | 需求核数，默认是 1。 |

### 配置文件

velocyto 软件需要对单细胞比对结果重新注释以评估 reads 代表的基因转录状态，区分未成熟 mRNA (unspliced) 和成熟 mRNA(spliced)。因此分析前需要先对上游比对产生的 bam 文件进行注释（可通过velocyto.prep.R 脚本完成）。

注释用参考文件（GTF）可在 ref 目录下找到，目前支持 human|mouse|rat，其他物种可按需自行添加和序列比对时用的GTF一致即可。

需要注意的是，目前 10x 流程比对结果没有代表 UMI 和 cell barcode （UB、CB）的标签，在使用本脚本注释时需要将参数 add\_tag 设置为 T 以自动加上。

注释后生成一个包含 4 个 assay 的 seurat 对象，可以直接用于之后的画图脚本。

画图脚本（velocyto.plot.R）最少需要两个 Seurat 对象，一个由velocyto.prep.R 生成，一个是经过降维需要标记速率的 Seurat 对象。注意两者的细胞名称需要检查一致。

### 脚本配置

run.sh：

#!bin/bash

#--bam /PROJ/development/xiezhuoming/proj/velocyto/input/demo/seekone\_demo/demo.bam \

#--filtered\_barcode /PROJ/development/xiezhuoming/proj/velocyto/input/demo/seekone\_demo/filtered\_barcodes.tsv.gz \

source /PROJ/development/xiezhuoming/soft/miniconda2/bin/activate /PROJ/development/xiezhuoming/soft/miniconda2/envs/velocyto

Rscript ./bin/velocyto.prep.R \

--bam /PROJ/development/xiezhuoming/proj/velocyto/input/demo/10xdemo/outs/possorted\_genome\_bam.bam \

--filtered\_barcode /PROJ/development/xiezhuoming/proj/velocyto/input/demo/10xdemo/outs/filtered\_feature\_bc\_matrix/barcodes.tsv.gz \

--species human \

--add\_tag F \

--outdir ./output \

--sample 10xdemo

plot.sh:

#!bin/bash

source /PROJ/development/xiezhuoming/soft/miniconda2/bin/activate /PROJ/development/xiezhuoming/soft/miniconda2/envs/velocyto

Rscript ./bin/velocyto.plot.R \

--velocyto\_rds /PROJ/development/xiezhuoming/proj/velocyto/output/demo/velocyto.rds \

--seurat\_obj2project /PROJ/development/xiezhuoming/proj/velocyto/input/demo/seekone\_demo/demo\_seurat.rds \

--outdir ./output \

--gene2show CCNL2,SSU72,errortest,GNB1

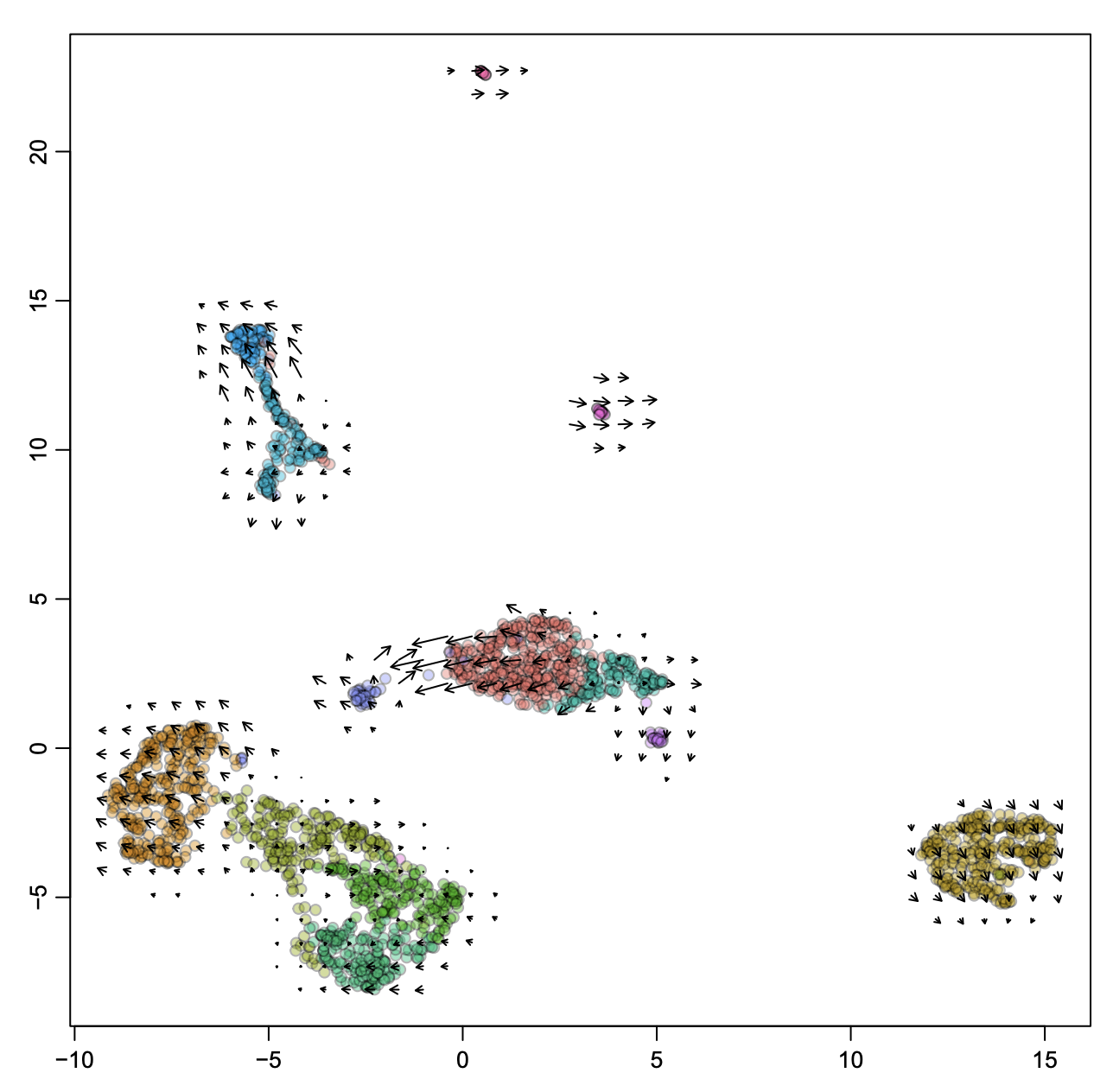
## 执行过程

先运行 run.sh 配置好相应文件后，再运行 plot.sh。

投递 qsub -cwd -l vf=50G,p=10 -V

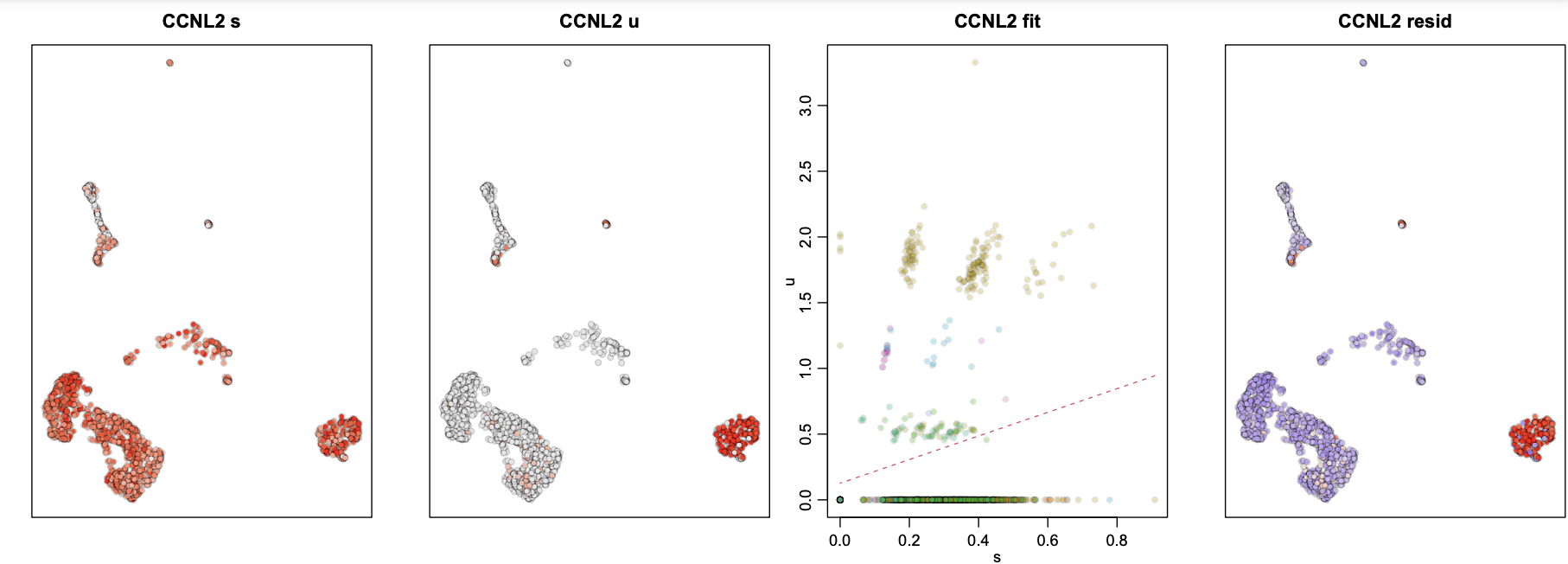
# 结果展示及解读

(1) velocyto\_projection.pdf：选定降维空间下的速率投射结果



箭头代表该处细胞速率预测的平均趋势。可以理解为细胞演化方向。

(2):<gene>.pdf： 单个基因的速率推断统计结果



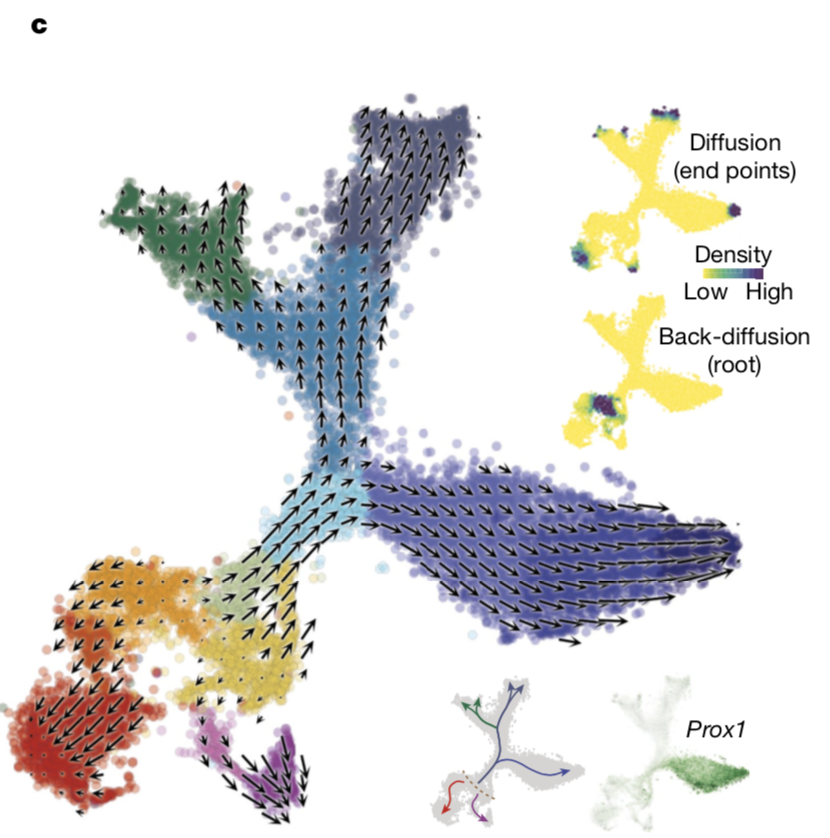
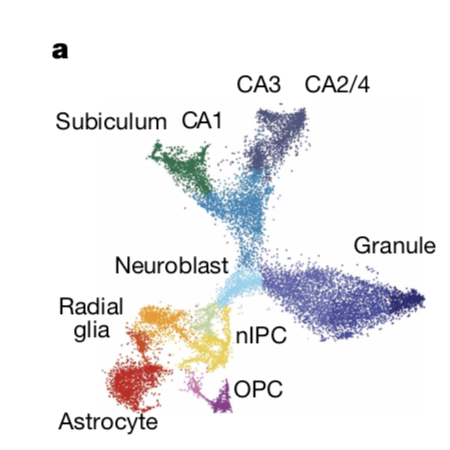
从左到右分别是选定基因的 splice reads（s）、unspliced reads(u) 统计结果、速率拟合结果、拟合结果残差分布。

（3）oldfit.rds

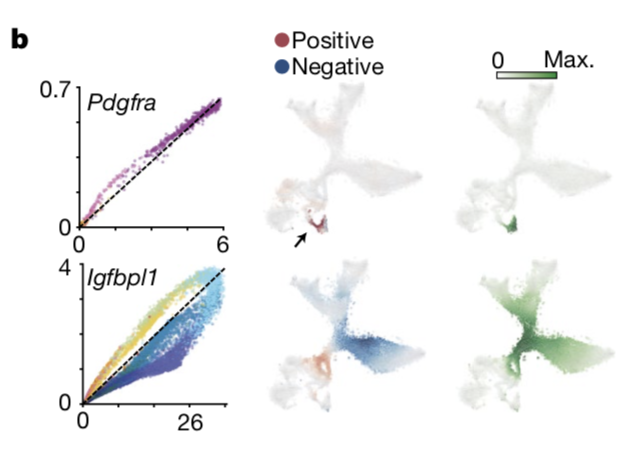
拟合 γ 值后的 velocyto 对象，可用于减少画图时间。

# 文献实例

在一个小鼠海马体细胞谱系研究中，RNA velocyto 交叉验证了细胞分类的正确性。



a 图是marker注释结果，c 图是 RNA velocyto 的速率投射结果，可以看出细胞是从 Neuroblast 向 CA 和 Granule 、preOPC 向 OPC 转化。



图为在 marker 基因层面的展示结果，左图是γ拟合结果，中图是u残差，右图是s表达量。

# 参考文献

1. La Manno, Gioele, et al. "RNA velocity of single cells."*Nature,*560.7719 (2018): 494-498.
2. Bergen, Volker, et al. "Generalizing RNA velocity to transient cell states through dynamical modeling."*Nature biotechnology,*38.12 (2020): 1408-1414.
3. Li, Tiejun, et al. "On the Mathematics of RNA Velocity I: Theoretical Analysis."*bioRxiv*(2020).
4. Qiu X, Zhang Y, Yang D, et al. Mapping vector field of single cells[J]. Biorxiv, 2021

# 问题反馈及建议

|  |  |  |  |
| --- | --- | --- | --- |
| 联系人 | 邮箱 | QQ | 备注 |
| 谢卓明 | xiezhuoming@seekgene.com |  |  |