# Single Cell RNA-seq

Guochao Li (liguochao@cncb.ac.cn)

Research Assistant from Jiang Lab

Beijing Institute of Genomics, Chinese Academy of Sciences

China National Center for Bioinformation

2025.07.11

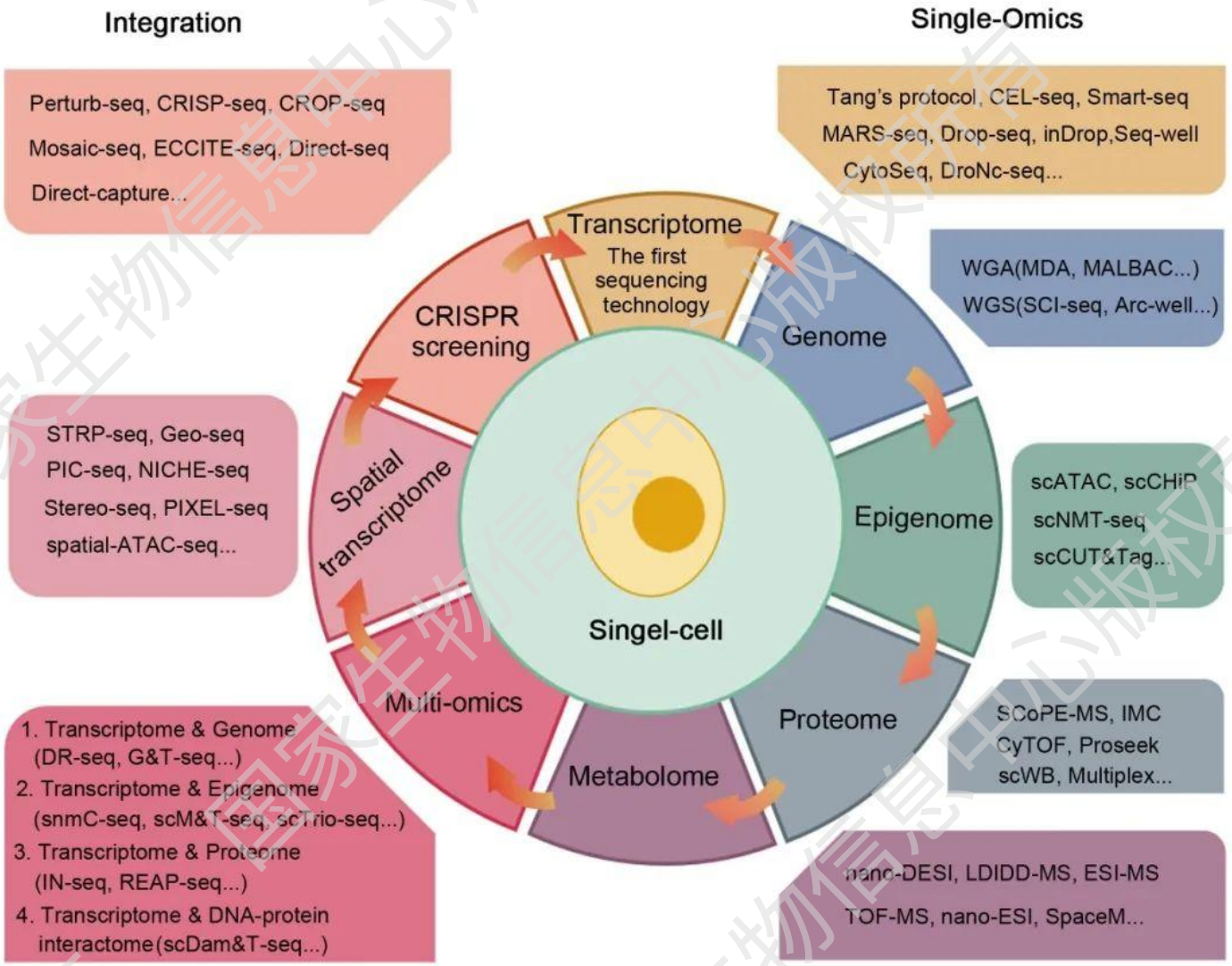
# Outline

1. Introduction of scRNA-seq
2. Analysis of scRNA-seq data 1：from 10X Genomics raw data (.fastq format) to cell x gene raw count matrix by cellranger count
3. Analysis of scRNA-seq data 2：Seurat standard pipeline for PBMC 3K
4. Analysis of scRNA-seq data 3：Further analysis by ggplot2, FindMarkers() and jjVolcano

# Outline

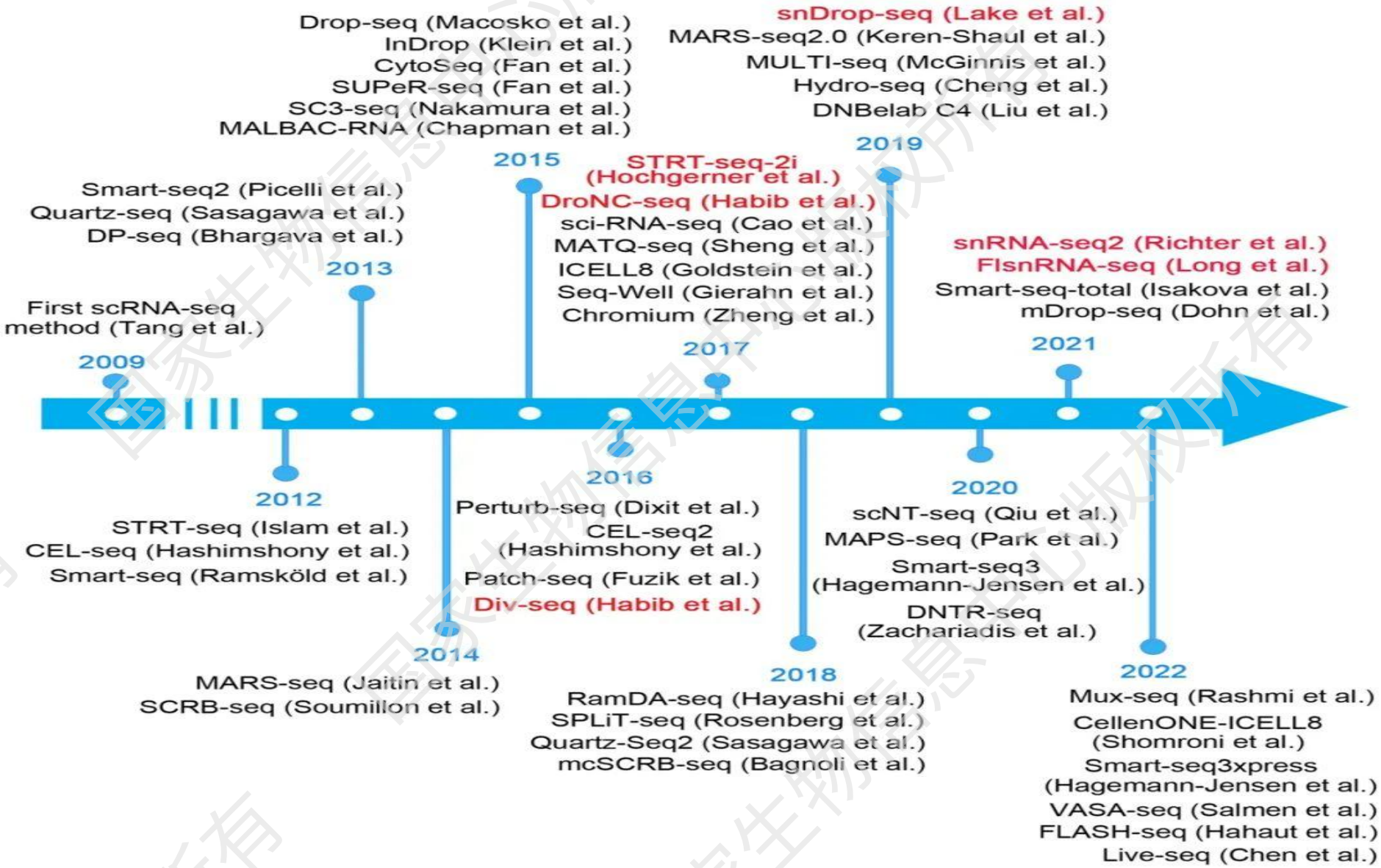
1. Introduction of scRNA-seq
2. Analysis of scRNA-seq data 1：from 10X Genomics raw data (.fastq format) to cell x gene raw count matrix by cellranger count
3. Analysis of scRNA-seq data 2：Seurat standard pipeline for PBMC 3K
4. Analysis of scRNA-seq data 3：Further analysis by ggplot2, FindMarkers() and jjVolcano

# Single cell technology  is becoming a powerful tool in many area

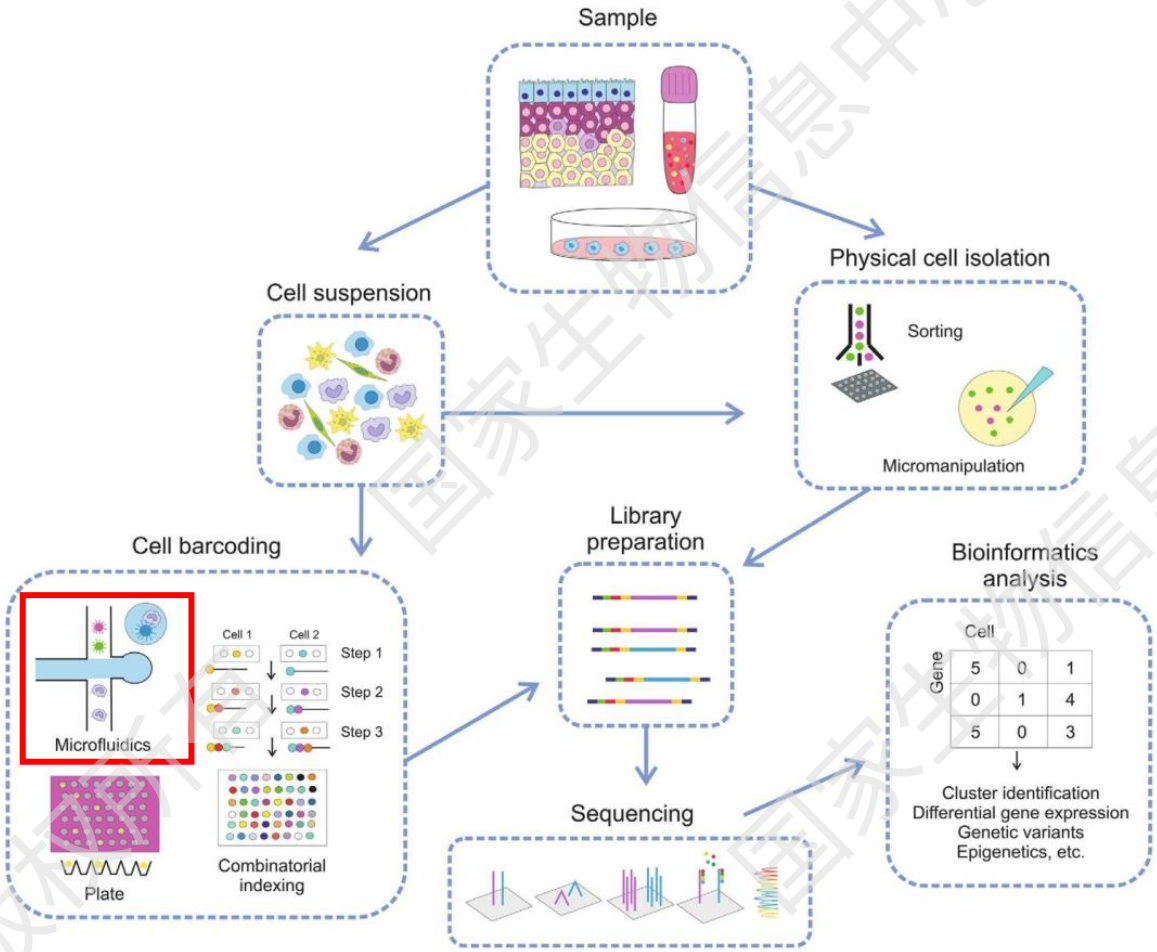


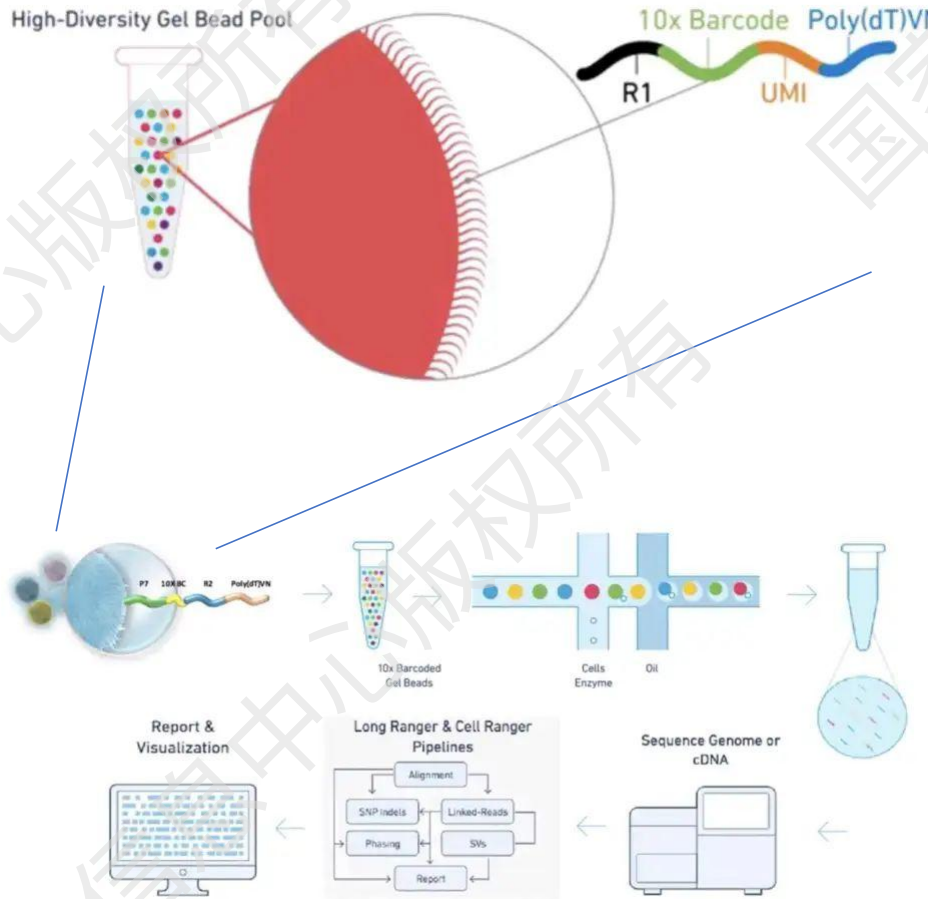

Sun, F., Li, H., Sun, D. *et al.* Single-cell omics: experimental workflow, data analyses and applications. *Sci. China Life Sci.* **68**, 5-102 (2025).

# Timeline of scRNA-seq



# scRNA-seq library preparation: droplet by 10X Genomics



# Outline

1. Introduction of scRNA-seq
2. Analysis of scRNA-seq data 1：from 10X Genomics raw data (.fastq format) to cell x gene raw count matrix by cellranger count
3. Analysis of scRNA-seq data 2：Seurat standard pipeline for PBMC 3K
4. Analysis of scRNA-seq data 3：Further analysis by ggplot2, FindMarkers() and jjVolcano

# 10X Genomics library preparation of 3’ scRNA

## Product support

Chromium

**Universal 3' Gene Expression**

3' gene expression profiling at scale with single cell resolution.

**Flex Gene Expression**

Fixed RNA Profiling assay for comprehensive probe-based gene expression profiling with single cell resolution.

**Epi ATAC**

Chromatin accessibilility profiling at the single cell level.

**Universal 5' Gene Expression**

5' gene expression alongside V(D)J repertoire profiling and antigen specificity of T and B cells.

**Epi Multiome ATAC + Gene Expression**

Combined profiling of 3' gene expression and chromatin accessibility from the same cell.

https://www.10xgenomics.com/support

# 10X Genomics library preparation of 3’ scRNA

Chromium Single Cell

## Universal 3' Gene Expression

This assay requires Chromium X/iX instrument Firmware v2.0 or newer. Refer to the Chromium X Series Firmware Update Procedure for details


### What is Universal 3' Gene Expression?

Chromium Single Cell Universal 3' Gene Expression provides single cell transcriptome 3' gene expression alongside the detection of surface protein expression in tens of thousands of cells.

Find user guides  Get started

# 10X Genomics library preparation of 3’ scRNA

## Documentation

We provide resources from protocols, user guides, and free-to-use analysis tools.

### Experimental Design & Planning

Plan ahead by referring to the list of equipment, kits, and reagents needed for Single Cell Gene Expression.

### Sample Prep

Prepare optimal single cell/nuclei suspension by following our Demonstrated Protocols.

### Instruments

Refer to instrument-specific documents for seamless instrument operation.

### Library Construction

Construct a sequencing-ready single cell library by following our step-by-step User Guide / Protocol.

### Targeted Gene Expression (Optional)

Use pre-designed or custom panels for targeted gene expression, followed by sequencing and data analysis.

### Sequencing

Sequence your single cell library as per our provided recommendations.

### Software Analysis

Analyze your single cell data using our pipelines, visualization, and Cloud Analysis tools.

# 10X Genomics scRNA rawdata in .fastq format

Path to rawdata/scRNA:

```
Sub01_S1_L001_R1_001.fastq.gz
Sub01_S1_L001_R2_001.fastq.gz
Sub01_S1_L002_R1_001.fastq.gz
Sub01_S1_L002_R2_001.fastq.gz
Sub01_S1_L003_R1_001.fastq.gz
Sub01_S1_L003_R2_001.fastq.gz
```

```
Sub02_S1_L001_R1_001.fastq.gz
Sub02_S1_L001_R2_001.fastq.gz
Sub02_S1_L002_R1_001.fastq.gz
Sub02_S1_L002_R2_001.fastq.gz
Sub02_S1_L003_R1_001.fastq.gz
Sub02_S1_L003_R2_001.fastq.gz
```

```
@A00511:134:HLLKJDSXX:4:1101:3360:1016 1:N:0:CCACTTAT
AATGGAAAGGGTTAATTTTGGCAATGGTCTAATATCTTATTTTTTTTTTTTTTTTTTTTTTTTTTTAAAATTTAATTTAAAAAAAAAAAAATAAAAAAAATAAAAAAAAAAAATTTTTTTGTTTTTTTTGGTTAATTGTTGGGGGTAAAAA
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFF,F,,,,,,F,,FFFFFFFFFFFFFFFFFFFF,,FFFF,,:F,,FFF,,,,,:FFF,F:,F,F,,F:FF,,,,,,:FFFFFFF,,,:,,:F,F,,,,FF:,:F,,F:F,F,:,FFF,,F::,,
@A00511:134:HLLKJDSXX:4:1101:6108:1016 1:N:0:CCACTTAT
GCCATGGCAGCACGAAAACTACATAGGGATAAGATTAAATTTTTTTTTTTTTTTTTTTTTAAAAAAAATGAAAAAGTAATATAAAAATAAAAAAAAAAAAAAAAAAAAAAAACCAAAAAAAATTAAAAAAAAATATAAAACCACACCCAACAAAA
+
:FF,:,,F:FF,FFFF:::FFF:F,:::F::F,F,FFFFFFFFFFFFFFF,,FFFF:F,,FFF,,:,F,,,FFF,,,F,F,F,,,,,,:F:FF:FFF,,F,,,,F,,F,,,,,,,,,:,FFF,,,:F,:FFF,,FF,F,,,,,,,F,FF
@A00511:134:HLLKJDSXX:4:1101:9796:1016 1:N:0:CCACTTAT
ATGGTTGAGCAGCCTCTCGTCTTCTTGGCTAATATATTATTTTTTTTTTTTTTTTTTTTTTTTTATTTTTAAAAAAATTTTTTTTTTTTTTTTTTTTAATTTTAAAAGATTTAAAAAAAAAAAACAAAAAACAAAAAAAACACCCCCACAAAAATTA
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFF,FFF:FFF,,:FFFFFFFFFFFFFFFFFFFF:,:,,,F,FFF,,F,FFFF:FFFFF:FFF,F,,,F,F,F:,,,,,,F,F,F,,,,,FF::FF,,,:,:FFF,,,,:FFFF,,F,,:,,:,,
```

# 10X Genomics scRNA rawdata in .fastq format

## Publications

Search abstracts, sample types, and more

| AREA OF INTEREST | SPECIES | TISSUE TYPE | JOURNAL | 10X GENOMICS PRODUCT |
|---|---|---|---|---|
| Cancer Research (1773) | Mouse (3531) | Tumor - solid (1187) | Cell Reports (358) | **Universal 3' Gene Expression** (7264) ✓ |
| Immunology (1551) | Human (3068) | Brain (1056) | Nature (293) | Universal 5' Gene Expression (1488) |
| Developmental Biology (1398) | Zebrafish (120) | Blood (707) | Cell (277) | Spatial Gene Expression for Fresh Frozen (713) |
| Neuroscience (1174) | Other - Mammal (81) | Lung (492) | Cell Stem Cell (208) | Linked-Reads Genomics (431) |
| Cell Biology (791) | Primate (70) | Bone Marrow (388) | Science Advances (166) | Epi ATAC (407) |
| Computational Method (400) | Plant (69) | Stem Cells (387) | | Epi Multiome ATAC (274) |
| Cell Atlas (372) | Drosophila (68) | Organoid (350) | | Spatial Gene Expression for FFPE (136) |
| | | | | In Situ Gene Expression (52) |
| | | | | Flex Gene Expression (27) |
| | | | | Single Cell CNV (25) |

# Cellranger count: rawdata → cell x gene raw count matrix

Example cmd in shell:

```
/data/biosoft/big/cellranger-9.0.1/cellranger count \
    --id test \
    --transcriptome /data/ref/refdata-gex-GRCh38-2020-A \
    --fastqs /data/biosoft/big/scRNA \
    --create-bam true \
    --localmem 30 \
    --localcores 10
```

Example output:

```
(base) [big@training test]$ ll outs
total 8652
drwxrwxr-x. 7 big big      74 Jun 26 23:02 analysis
-rw-rw-r--. 2 big big  858097 Jun 26 23:02 cloupe.cloupe
drwxrwxr-x. 2 big big      73 Jun 26 23:01 filtered_feature_bc_matrix
-rw-rw-r--. 1 big big  352493 Jun 26 23:01 filtered_feature_bc_matrix.h5
-rw-rw-r--. 1 big big     640 Jun 26 23:02 metrics_summary.csv
-rw-rw-r--. 1 big big  896972 Jun 26 23:01 molecule_info.h5
-rw-rw-r--. 1 big big  307802 Jun 26 23:01 possorted_genome_bam.bam
-rw-rw-r--. 1 big big 1488208 Jun 26 23:01 possorted_genome_bam.bam.bai
drwxrwxr-x. 2 big big      73 Jun 26 23:01 raw_feature_bc_matrix
-rw-rw-r--. 1 big big  502438 Jun 26 23:01 raw_feature_bc_matrix.h5
-rw-rw-r--. 1 big big 4432121 Jun 26 23:02 web_summary.html
```

https://www.10xgenomics.com/support/software/cell-ranger/latest

# Content of three output files

```
(base) [big@training test]$ cd outs/filtered_feature_bc_matrix/
(base) [big@training filtered_feature_bc_matrix]$ ls
barcodes.tsv.gz  features.tsv.gz  matrix.mtx.gz
(base) [big@training filtered_feature_bc_matrix]$ zcat barcodes.tsv.gz | head
AAACGAACAAGCTACT-1
AAACGCTAGGAGATAG-1
AAAGAACAGAATCGAT-1
AAAGGATCAATGGGTG-1
AAAGGATCAGATCCTA-1
AAAGGATTCCTAGCGG-1
AAAGGGCGTTTGAAAG-1
AAAGGTAGTCTTACTT-1
AAAGTCCCAACAGTGG-1
AAAGTCCCACAACCGC-1
(base) [big@training filtered_feature_bc_matrix]$ zcat features.tsv.gz | head
ENSG00000243485 MIR1302-2HG     Gene Expression
ENSG00000237613 FAM138A Gene Expression
ENSG00000186092 OR4F5   Gene Expression
ENSG00000238009 AL627309.1      Gene Expression
ENSG00000239945 AL627309.3      Gene Expression
ENSG00000239906 AL627309.2      Gene Expression
ENSG00000241860 AL627309.5      Gene Expression
ENSG00000241599 AL627309.4      Gene Expression
ENSG00000286448 AP006222.2      Gene Expression
ENSG00000236601 AL732372.1      Gene Expression
(base) [big@training filtered_feature_bc_matrix]$ zcat matrix.mtx.gz | head
%%MatrixMarket matrix coordinate integer general
%metadata_json: {"software_version": "cellranger-9.0.1", "format_version": 2}
36601 1359 1487
34658 1 2
31576 2 1
1540 3 1
36569 4 1
35379 5 1
5887 6 1
25618 7 1
```

# Outline

1. Introduction of scRNA-seq
2. Analysis of scRNA-seq data 1：from 10X Genomics raw data (.fastq format) to cell x gene raw count matrix by cellranger count
3. Analysis of scRNA-seq data 2：Seurat standard pipeline for PBMC 3K
4. Analysis of scRNA-seq data 3：Further analysis by ggplot2, FindMarkers() and jjVolcano

# Seurat in action

Seurat 5.2.0 | Install | Get started | Vignettes | Extensions | FAQ | News | Reference | Archive


## Seurat v5

We are excited to release Seurat v5! To install, please follow the instructions in our install page. This update brings the following new features and functionality:

- **Integrative multimodal analysis:** The cellular transcriptome is just one aspect of cellular identity, and recent technologies enable routine profiling of chromatin accessibility, histone modifications, and protein levels from single cells. In Seurat v5, we introduce 'bridge integration', a statistical method to integrate experiments measuring different modalities (i.e. separate scRNA-seq and scATAC-seq datasets), using a separate multiomic dataset as a molecular 'bridge'. For example, we demonstrate how to map scATAC-seq datasets onto scRNA-seq datasets, to assist users in interpreting and annotating data from new modalities.

  We recognize that while the goal of matching shared cell types across datasets may be important for many problems, users may also be concerned about which method to use, or that integration could result in a loss of biological resolution. In Seurat v5, we also introduce flexible and streamlined workflows for the integration of multiple scRNA-seq datasets. This makes it easier to explore the results of different integration methods, and to compare these results to a workflow that excludes integration steps.

### Links

View on CRAN

Browse source code

Report a bug

### License

Full license

MIT + file LICENSE

### Community

Code of conduct

### Citation

Citing Seurat

### Developers

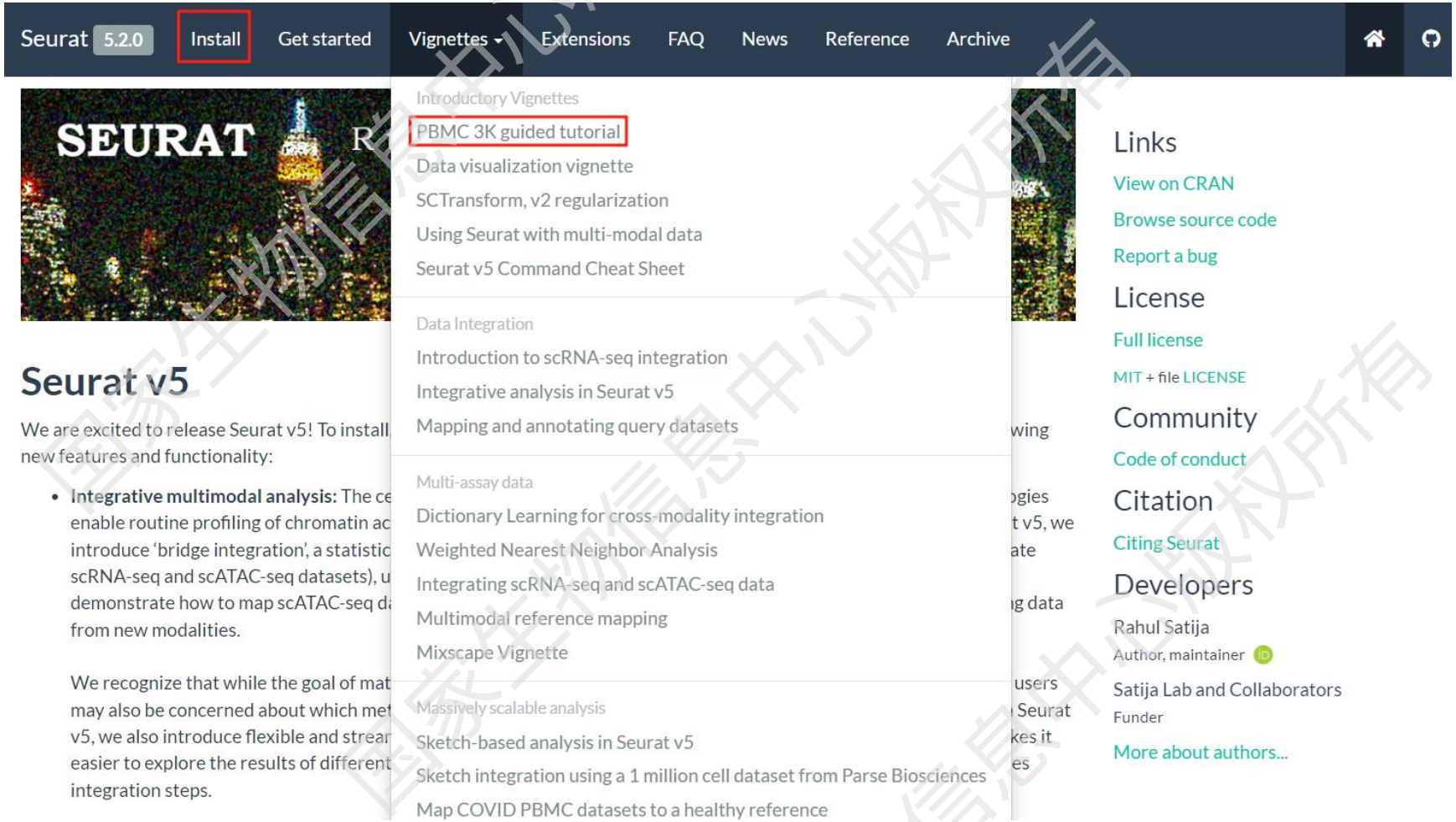
Rahul Satija
Author, maintainer

Satija Lab and Collaborators
Funder

More about authors...

https://satijalab.org/seurat

# Seurat in action




https://satijalab.org/seurat

# Outline

1. Introduction of scRNA-seq
2. Analysis of scRNA-seq data 1：from 10X Genomics raw data (.fastq format) to cell x gene raw count matrix by cellranger count
3. Analysis of scRNA-seq data 2：Seurat standard pipeline for PBMC 3K
4. Analysis of scRNA-seq data 3：Further analysis by ggplot2, FindMarkers() and jjVolcano

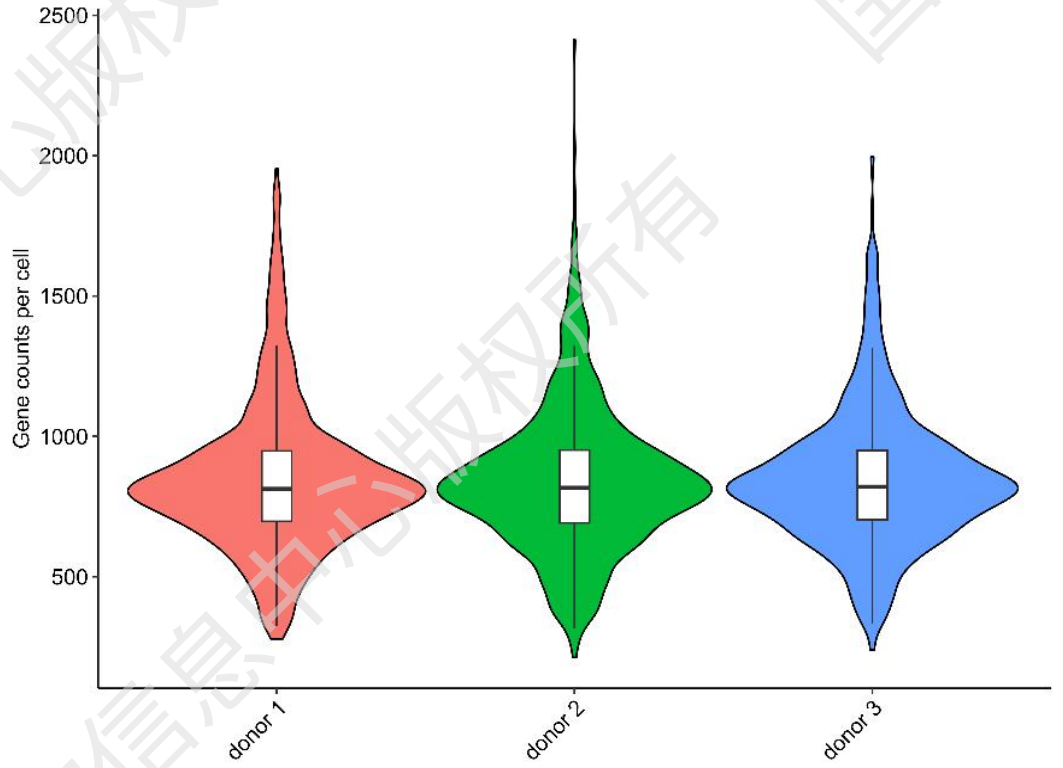
# violin plot of gene counts per cell for each donor

常用质控图：此处假设PBMC 3K个细胞来源于3个donor，展示不同donor各自的gene counts per cell distribution

```
library(Seurat)
library(ggplot2)
library(data.table)
library(ggpubr)

pbmc = readRDS("./pbmc3k.rds")
head(pbmc@meta.data)
pbmc@meta.data$cell_type = Idents(pbmc)
pbmc@meta.data$donor = rep(paste0("donor ",c(1:3)), 863)
head(pbmc@meta.data)
metadata = as.data.table(pbmc@meta.data, keep.rownames = "BARCODE")

# violin plot for gene counts distribution
p = ggplot(data=metadata,aes(x=donor,y=nFeature_RNA,fill=donor)) +
  geom_violin(color=I("black"), width=1) +
  geom_boxplot(fill=I("white"), width = 0.1, outlier.shape = NA) +
  xlab(element_blank()) +
  ylab("Gene counts per cell") +
  theme_pubr() +
  theme(axis.text.x = element_text(angle=45, hjust=1),
        legend.position = "none")
ggsave("ggplot2.violin.gene.png", plot=p, dpi=600, height = 6, width = 8)
```



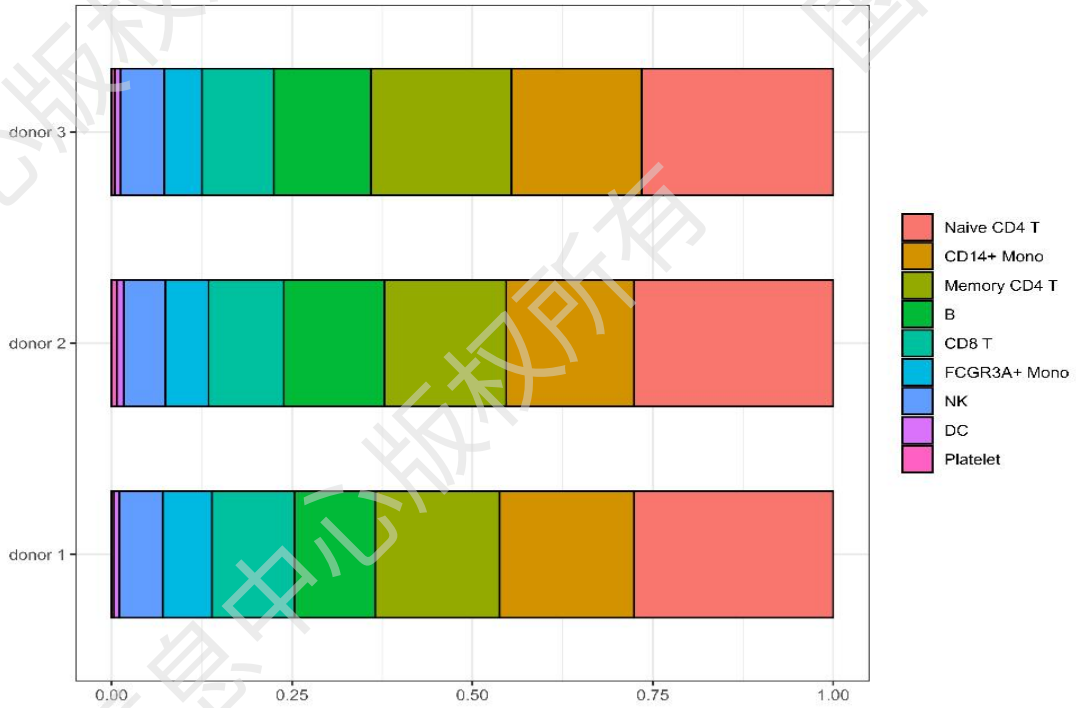
# barplot for cell type ratio change among donors

常用图：对比不同donor（或不同group）的cell type ratio 变化情况

```
library(Seurat)
library(ggplot2)
library(data.table)
library(ggpubr)

pbmc = readRDS("./pbmc3k.rds")
head(pbmc@meta.data)
pbmc@meta.data$cell_type = Idents(pbmc)
pbmc@meta.data$donor = rep(paste0("donor ",c(1:3)), 863)
head(pbmc@meta.data)
metadata = as.data.table(pbmc@meta.data, keep.rownames = "BARCODE")

# barplot for cell type ratio
j = metadata[, .N, by=.(donor, cell_type)]
for (donor_tmp in unique(j$donor)) {
  j[donor==donor_tmp, percent := N/sum(N)]
}
p = ggplot(data=j, aes(x=percent,y=donor,fill=cell_type)) +
  geom_bar(stat="identity", color="black", width = 0.6) +
  ylab(element_blank()) +
  xlab(element_blank()) +
  theme_bw() +
  theme(legend.title = element_blank(), legend.position = "right")
ggsave(filename="./ggplot2.barplot.celltype_ratio.png", plot=p, dpi=600, width=8, height=6)
```



# case vs control DEG in each cell type

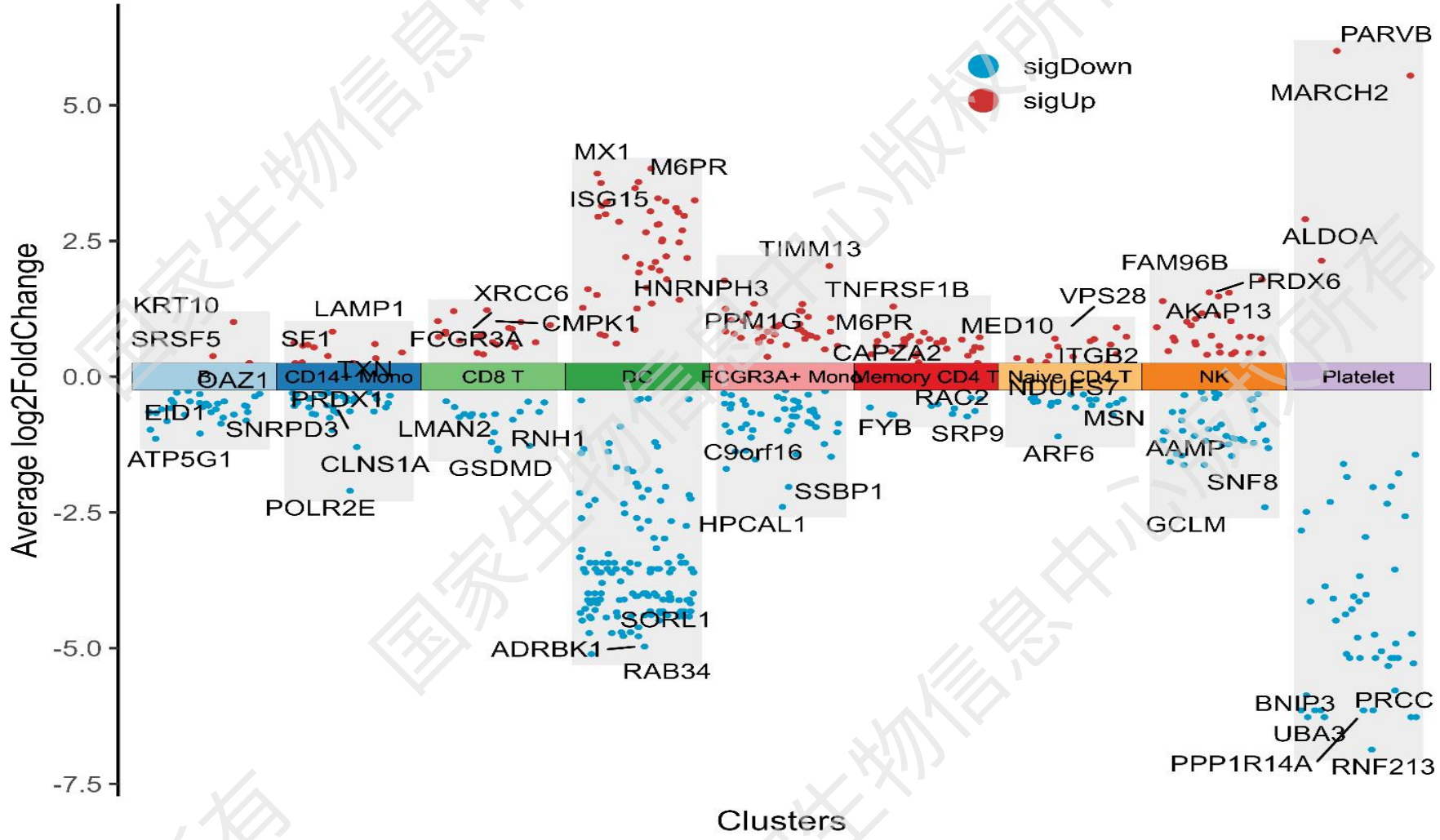
灵活运用FindMarkers函数，得到每个cell type中的case vs control差异表达基因，并使用jjVolcano进行可视化

```
library(scRNAtoolVis)

# visualization of DEG (differential expressed gene) in each cell type
pbmc@meta.data$group = rep(c("control", "case", "case"), 863)
head(pbmc@meta.data)
markers = data.table()
for (cell_type_tmp in unique(pbmc@meta.data$cell_type)) {
  # Set logfc.threshold = 0 to save all genes, not only significantly DEG, for plot.
  tmp = FindMarkers(pbmc, ident.1 = "case", ident.2 = "control", group.by = 'group', logfc.threshold = 0, min.pct
= 0.25, subset.ident = cell_type_tmp)
  setDT(tmp, keep.rownames="gene")
  tmp[,cluster:=cell_type_tmp]
  markers = rbind(markers, tmp)
}
fwrite(markers, file="genes_exp_in_each_cell_type.txt", sep="\t")
p = jjVolcano(diffData = markers, topGeneN = 3)
ggsave(filename="./ggplot2.jjVolcano.DEG_in_each_cell_type.png", plot=p, dpi=600, width=8, height=6)
```

# case vs control DEG in each cell type

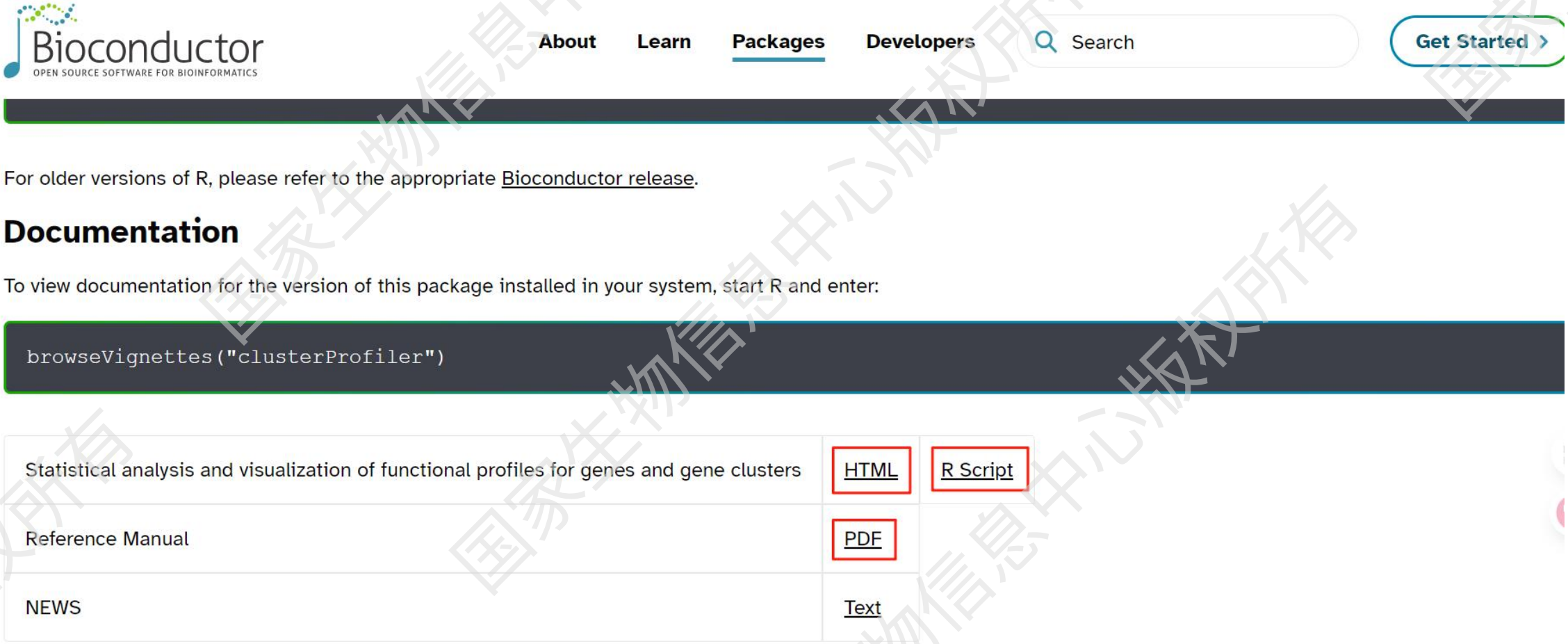
灵活运用FindMarkers函数，得到每个cell type中的case vs control差异表达基因，并使用jjVolcano进行可视化



# Using Bioconductor to learn clusterProfiler (KEGG/GO/GSEA ...)


https://www.bioconductor.org/

# Using Bioconductor to learn clusterProfiler (KEGG/GO/GSEA ...)



# Using Bioconductor to learn clusterProfiler (KEGG/GO/GSEA ...)

## Visualization

- barplot
- cnetplot
- dotplot
- emapplot
- gseaplot
- goplot
- upsetplot

## Vignette

Please go to https://yulab-smu.github.io/clusterProfiler-book/ for the full vignette.

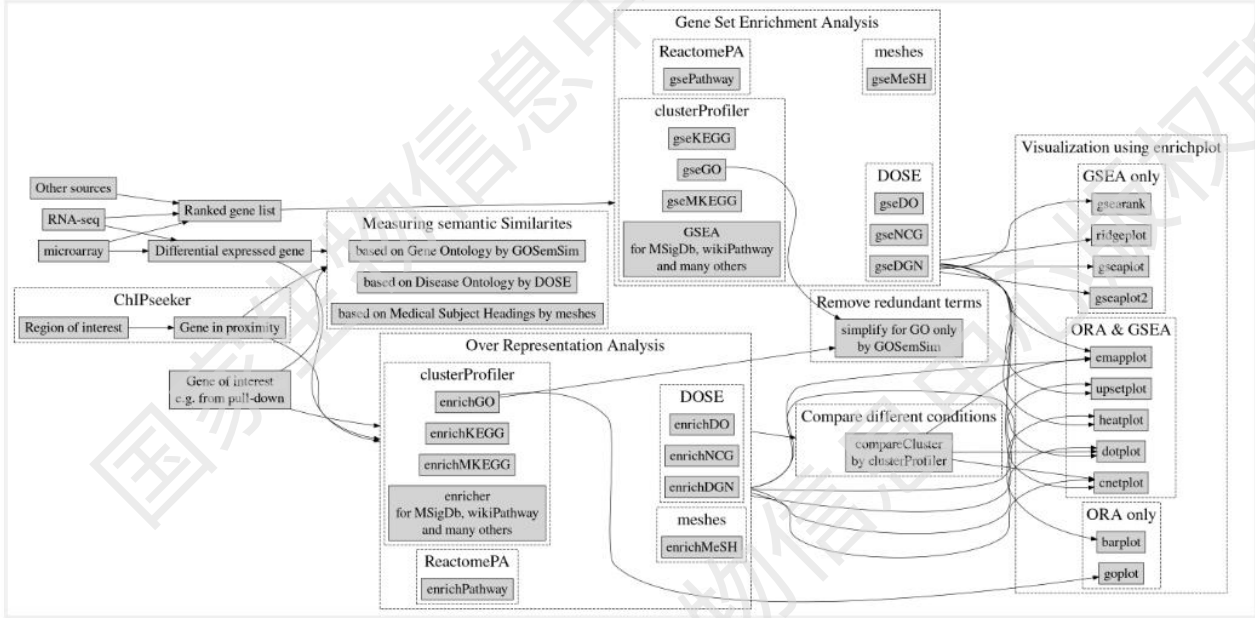
# Using Bioconductor to learn clusterProfiler (KEGG/GO/GSEA ...)



## Introduction

ChIPseeker 1.30.3 | clusterProfiler 4.3.1.900 | DOSE 3.21.1.9002 | enrichplot 1.15.3 | GOSemSim 2.21.1 | meshes 1.20.0 | ReactomePA 1.38.0

hello




On this page

# Thank you!

http://jianglab.group