

# 单细胞测序数据挖掘



# 主要内容

- 简介
- 数据下载
- 数据整理
- 质控和数据过滤
- PCA分析
- TSNE分析
- Marker基因
- 注释细胞类型
- 细胞轨迹分析
- GO富集分析和圈图
- KEGG富集分析和圈图

# 单细胞测序

单细胞测序是指在单个细胞水平上进行测序。单细胞转录组测序（single cell RNA Seq, scRNA-seq）是指对于单个细胞水平上将mRNA反转录扩增后进行高通量测序的技术。

单细胞测序通过在单个细胞水平上进行测序，解决了用组织样本无法获得不同细胞间的异质性信息或样本量太少无法进行常规测序的难题，为科学家研究单个细胞的行为、机制等提供了新的方向。

单细胞基因组测序主要包括四个步骤：单细胞分离→扩增→高通量测序→数据分析。

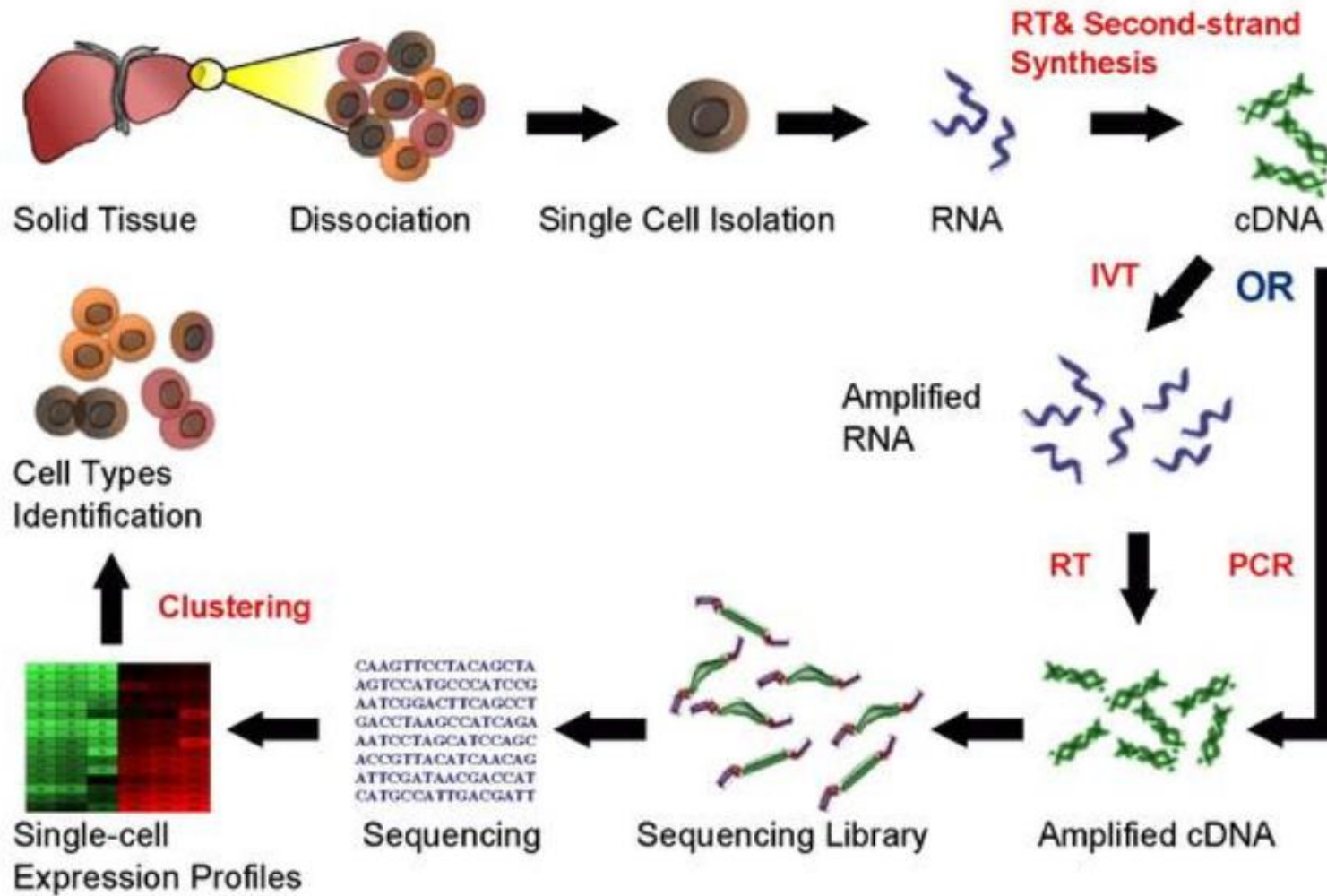
# 单细胞测序

目前常规的测序主要是数百万甚至更多细胞的混合DNA样本。这种方法能够得到基因表达信息，但是对其进行研究得到的结果只是一群细胞中信号的平均值，或者只代表其中占优势数量的细胞信息，单个细胞独有的特性被忽视。

单细胞RNA-seq能够独立地提供每个细胞的RNA表达谱，并鉴定异质细胞群中的稀有细胞。尽管肿瘤异质性可归因于累积突变，但即使是遗传上相同的细胞在相同环境下也可能表现出基因和蛋白表达水平的差异，单细胞RNA-seq就能够发现这些稀有个体。比如在肿瘤组织中，肿块中心的细胞，肿块周围的细胞，淋巴转移灶的细胞，以及远端转移的细胞，其基因组和转录组等遗传信息，是存在差异的。

# 单细胞测序

## Single Cell RNA Sequencing Workflow



# 数据下载

NCBI Resources How To

[Sign in to NCBI](#)

[GEO Home](#)

[Documentation](#)

[Query & Browse](#)

[Email GEO](#)

## Gene Expression Omnibus



GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Search

### Getting Started

[Overview](#)

[FAQ](#)

[About GEO DataSets](#)

[About GEO Profiles](#)

[About GEO2R Analysis](#)

[How to Construct a Query](#)

[How to Download Data](#)

### Tools

[Search for Studies at GEO DataSets](#)

[Search for Gene Expression at GEO Profiles](#)

[Search GEO Documentation](#)

[Analyze a Study with GEO2R](#)

[Studies with Genome Data Viewer Tracks](#)

[Programmatic Access](#)

[FTP Site](#)

### Browse Content



生信自学网  
[www.biowolf.cn](http://www.biowolf.cn)

[Repository Browser](#)

DataSets: 4348

Series:  114163

Platforms: 19798

Samples: 3097980

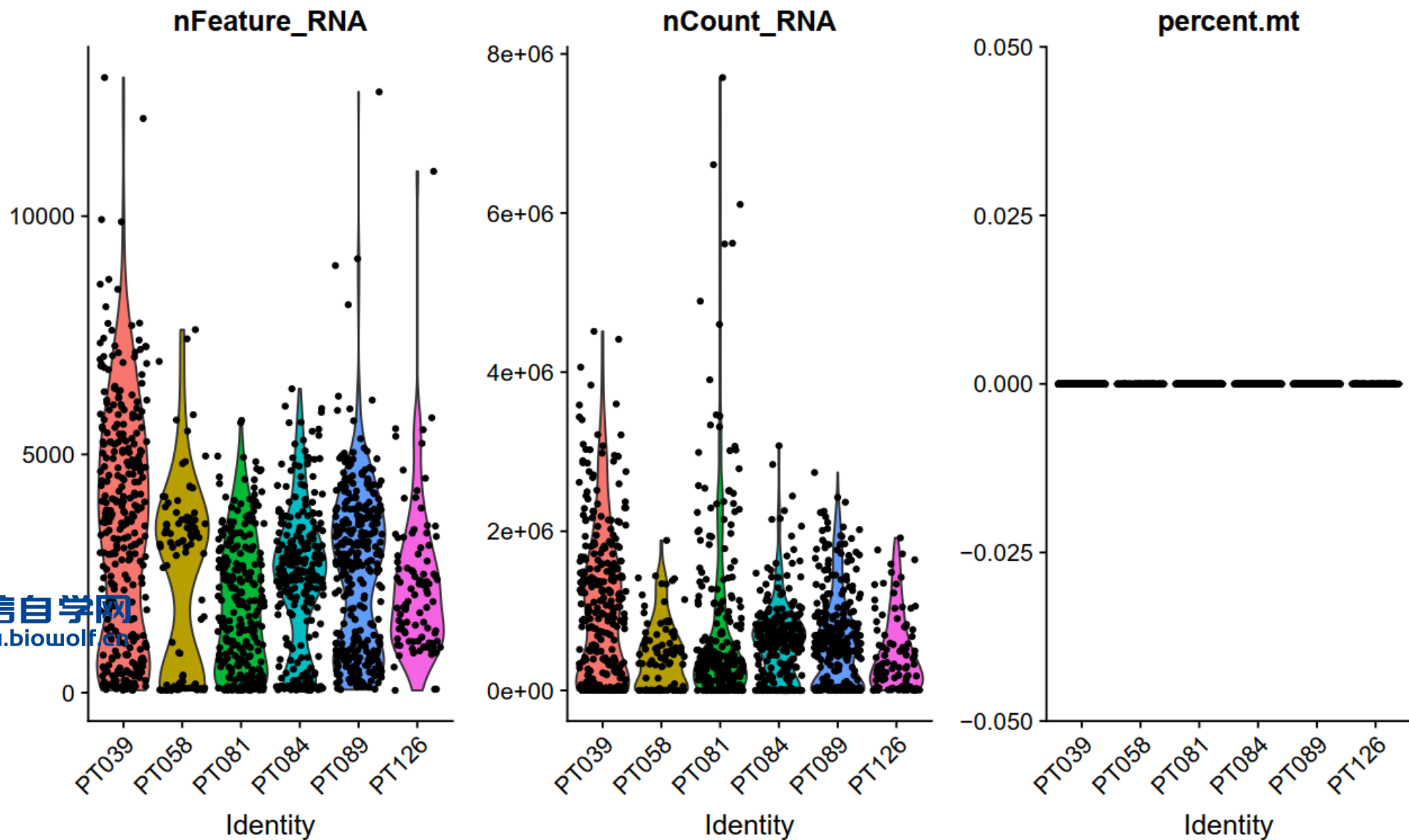
# 数据整理

id	PT089_P1_A01	PT089_P1_A02	PT089_P1_A03	PT089_P1_A04
AAGAB	0	20	450	0
AAK1	5.85	10.08	13.65	4.41
AAMDC	226	0	0	0
AAMP	911	9	0	1
AANAT	0	0	0	0
AAR2	0	0	0	0
AARD	1.15	1.01	0	0
AARS	0	0	0	0
AARS2	1.15	6.14	0	0
AARSD1	0	0	0	0



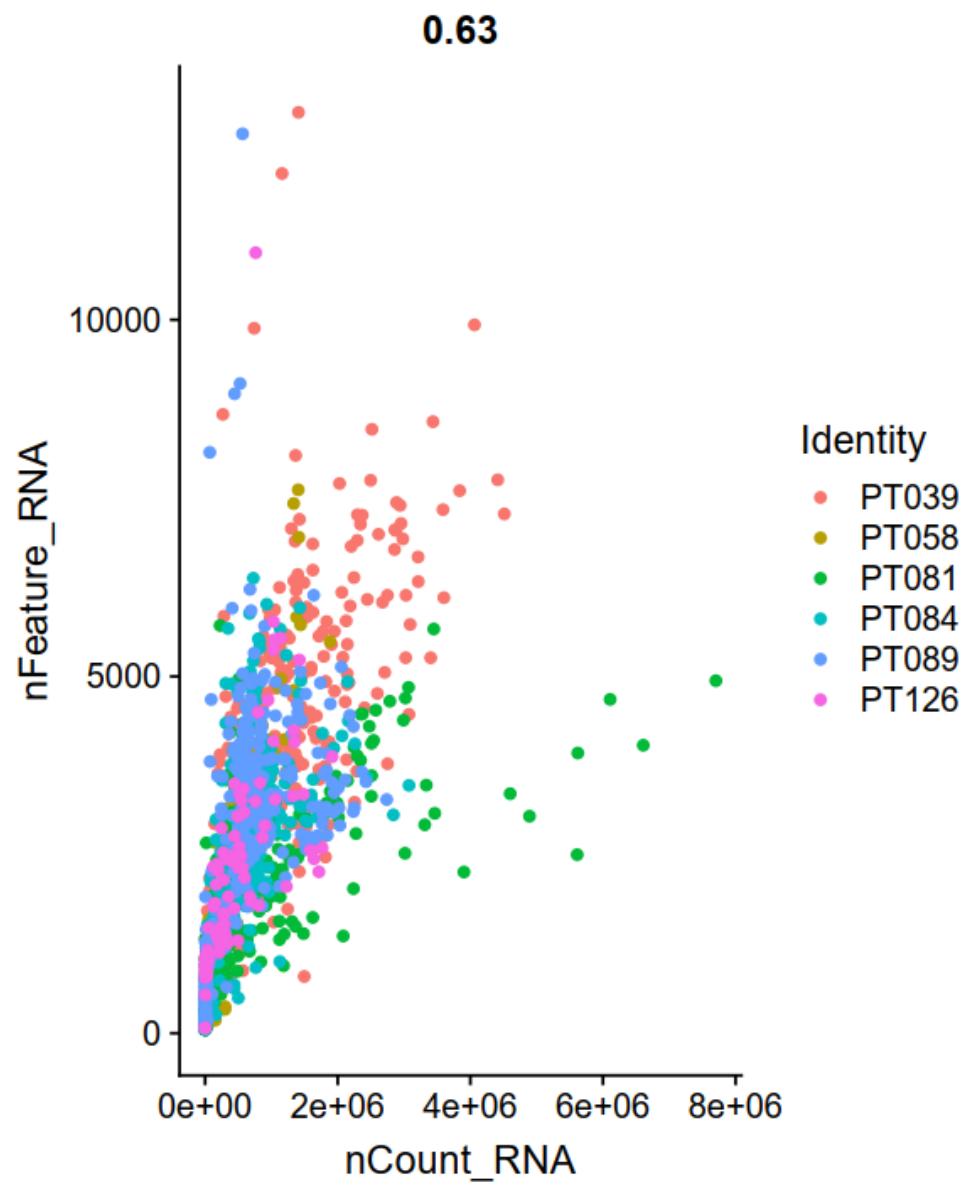
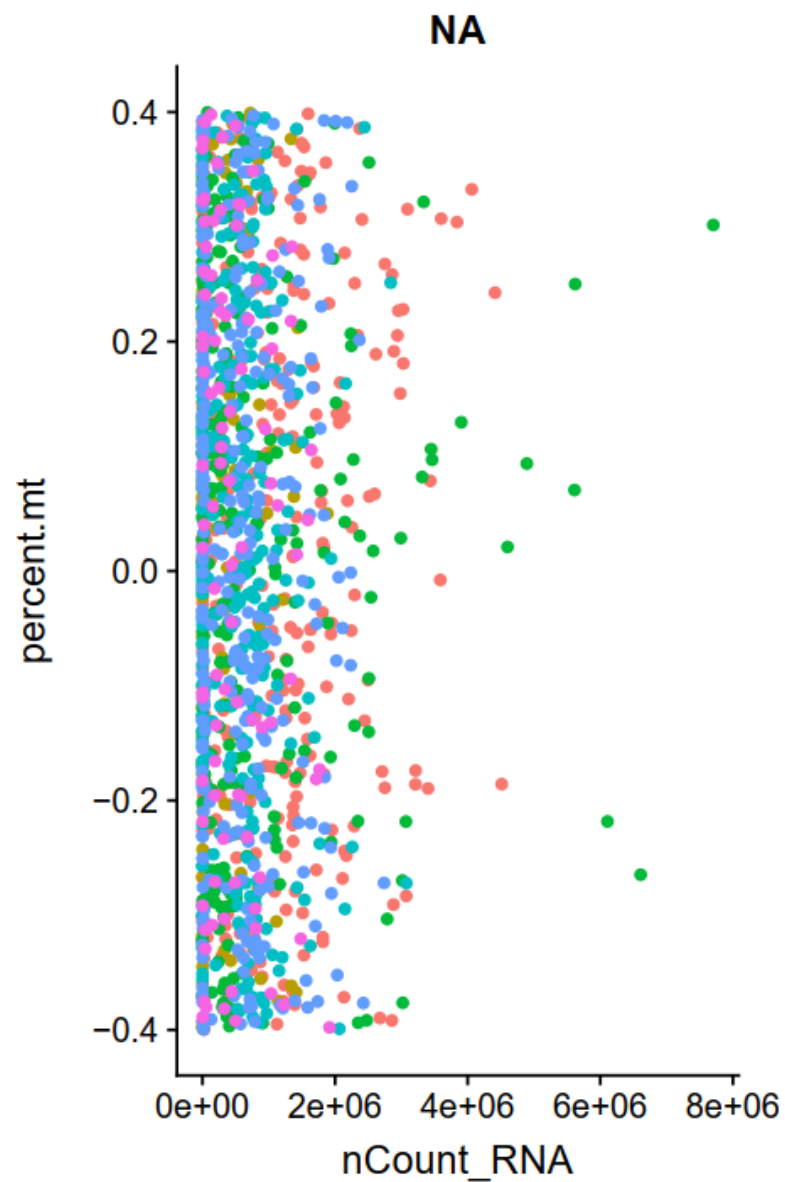
生信自学网  
www.biowolf.cn

# 质控和数据过滤

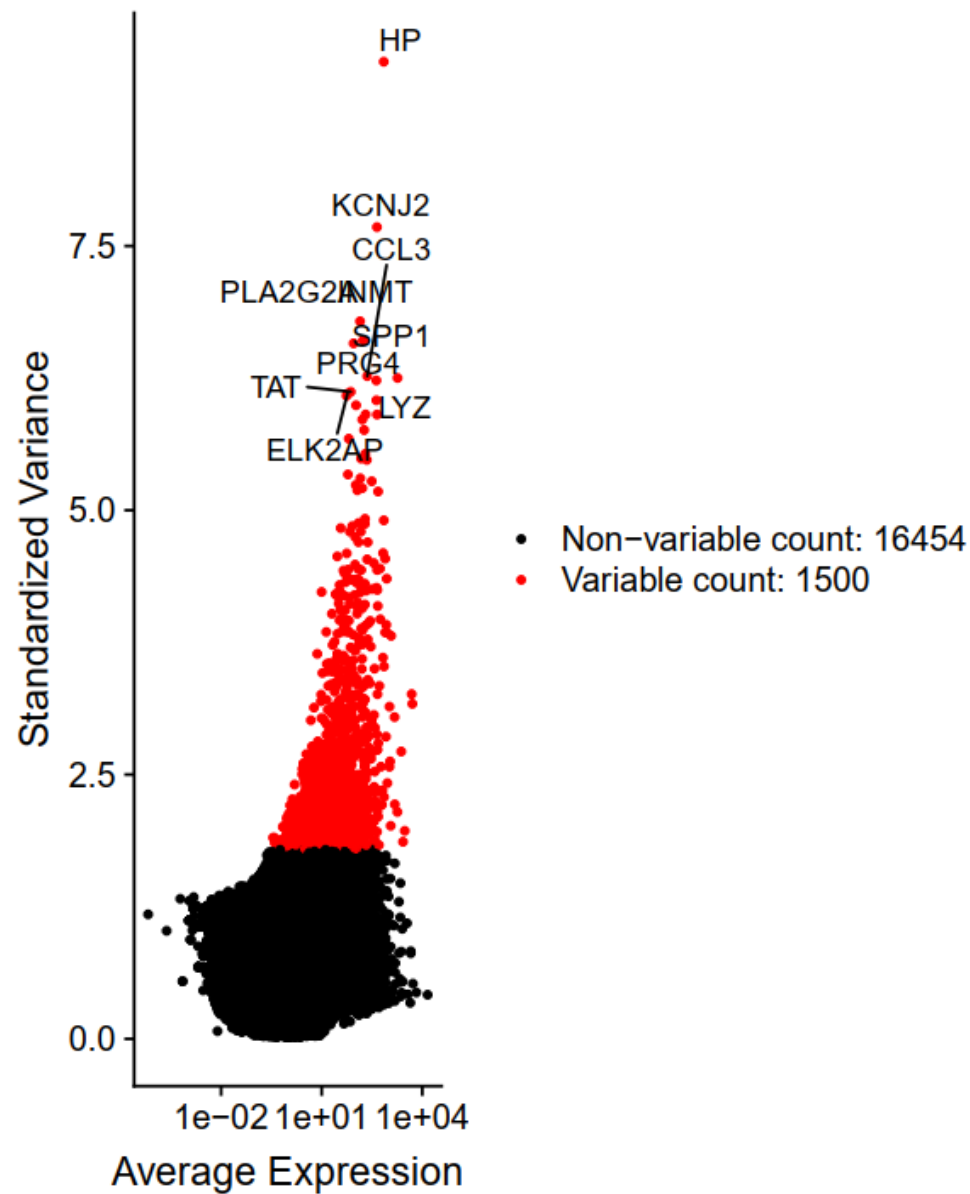
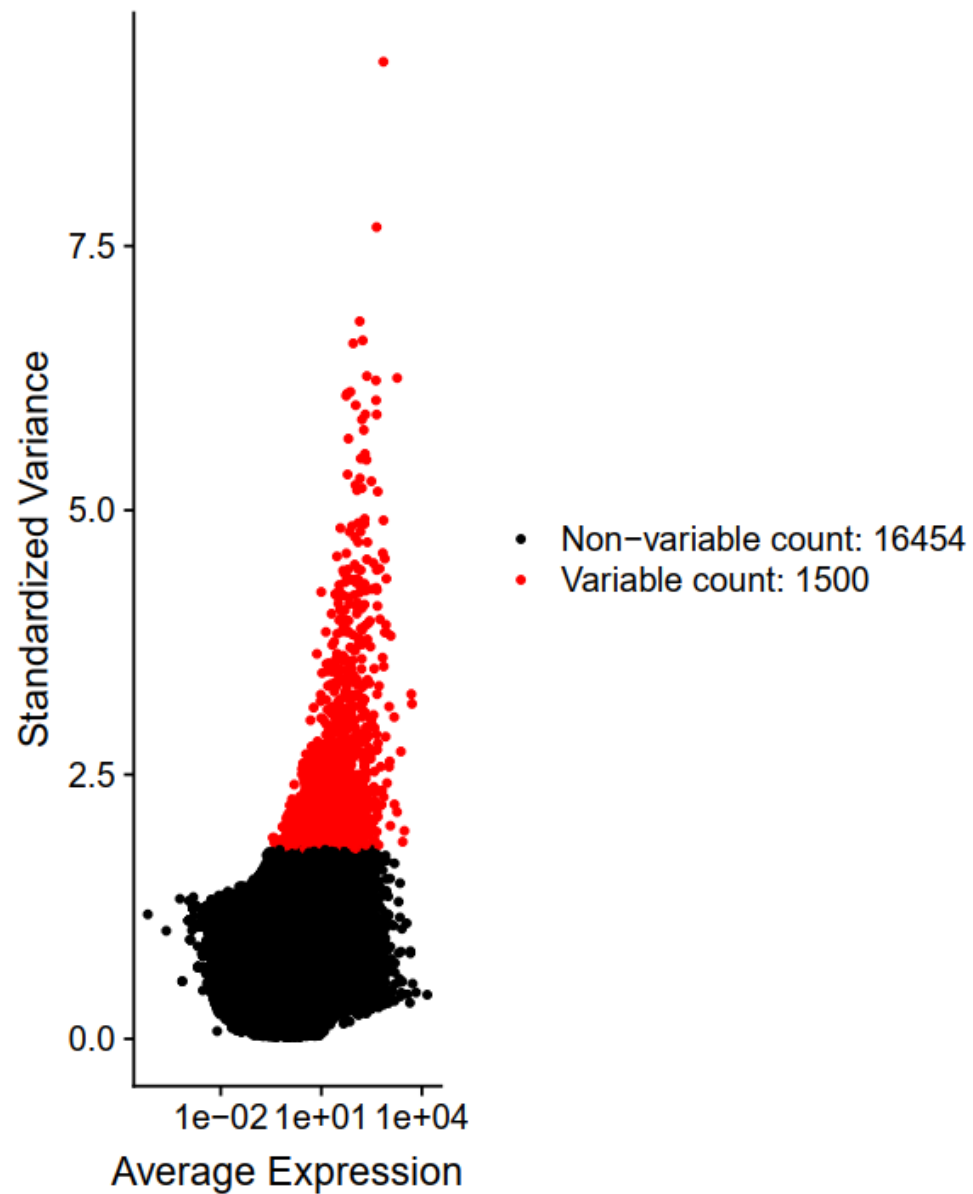




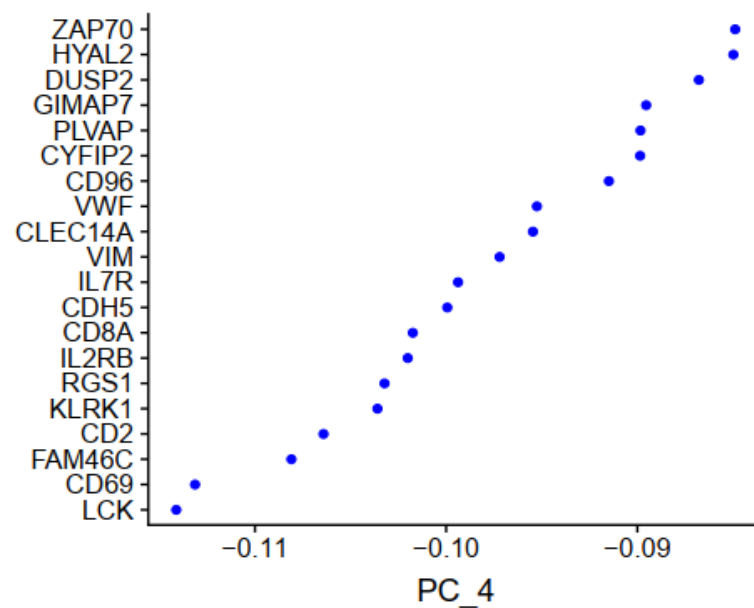
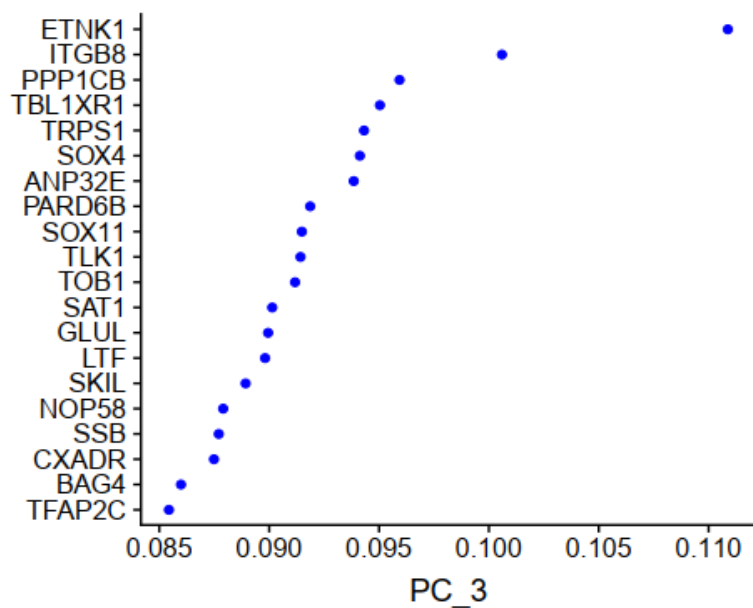
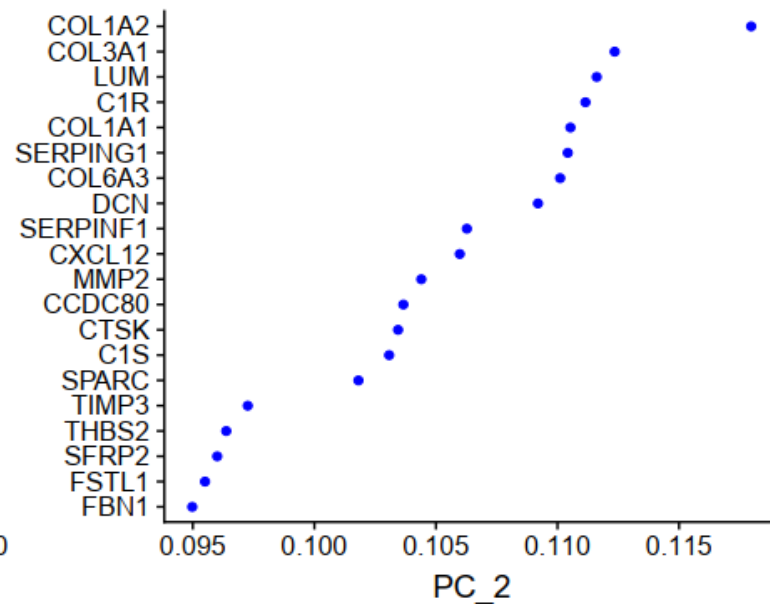
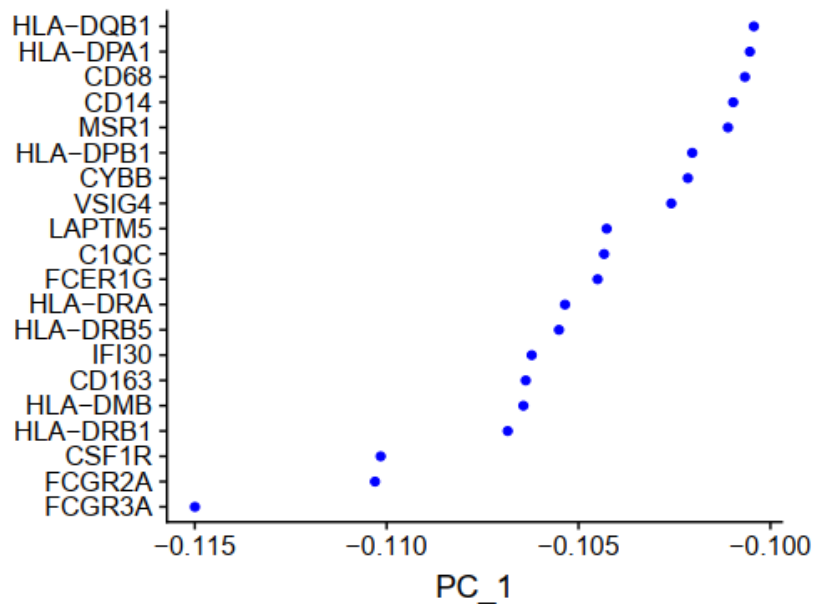
# 质控和数据过滤



# 质控和数据过滤

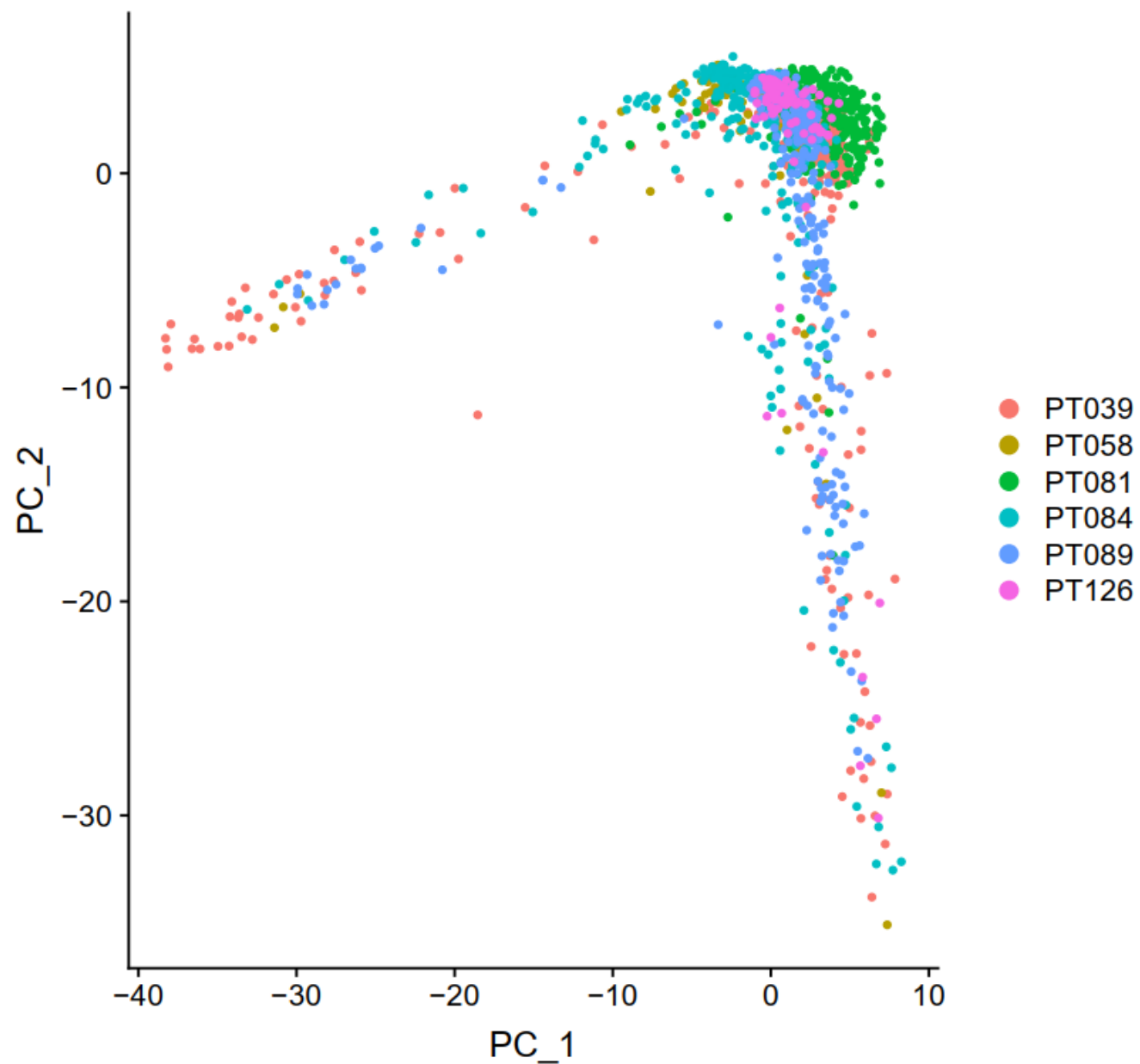


# PCA主成分分析



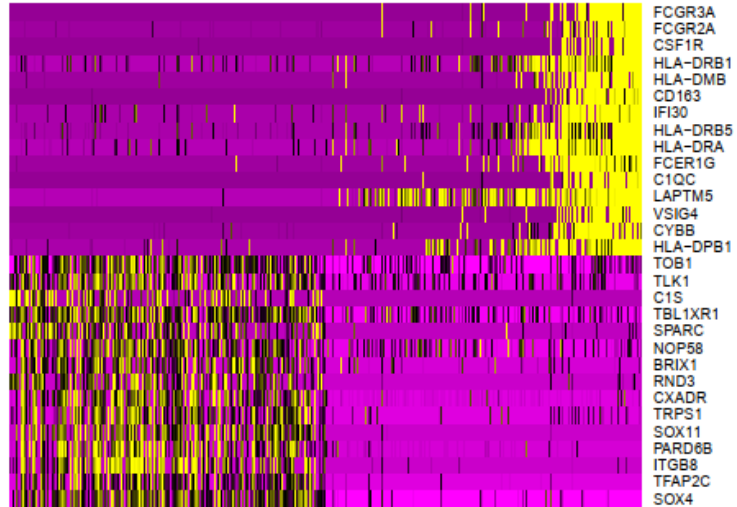


# PCA主成分分析

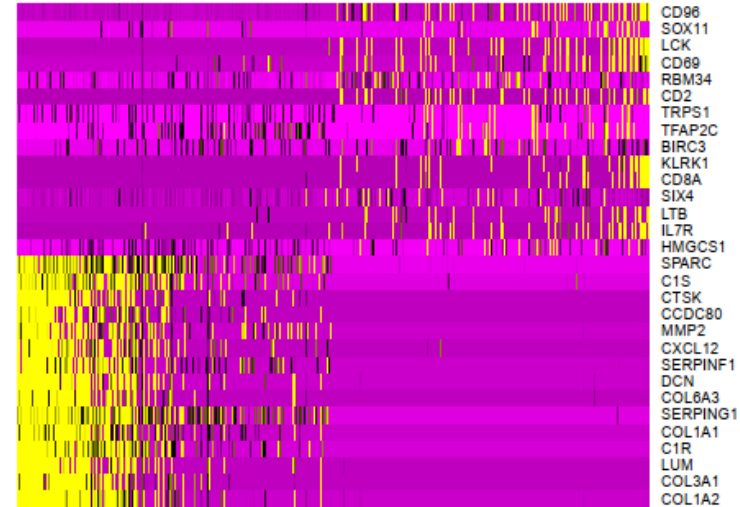


# PCA主成分分析

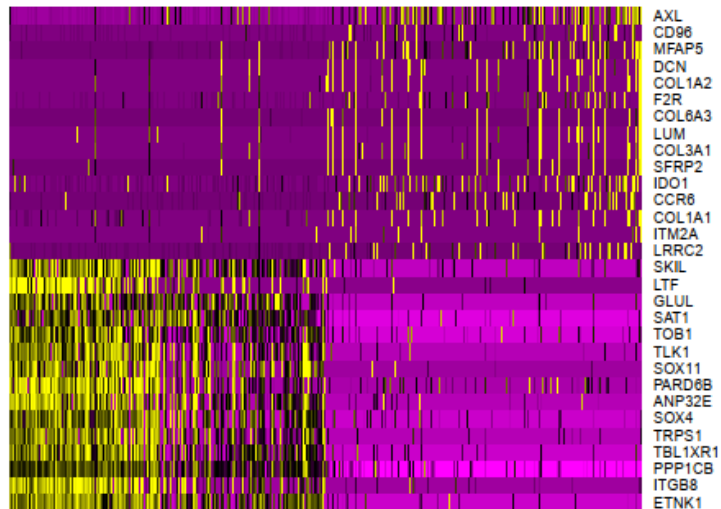
PC\_1



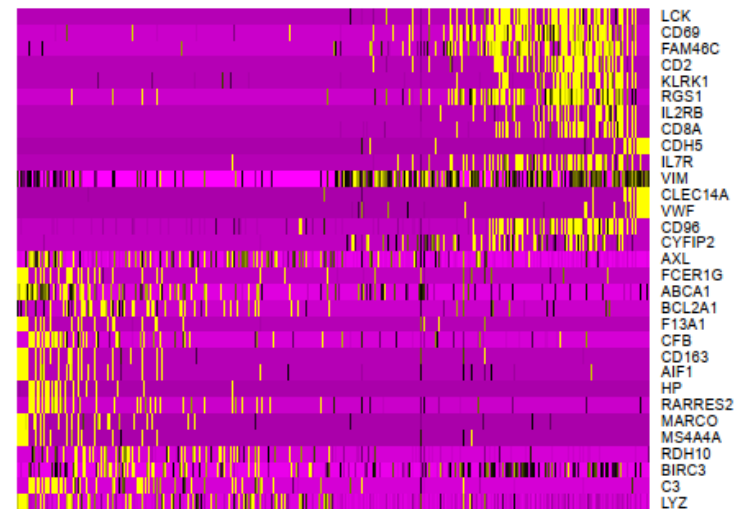
PC\_2



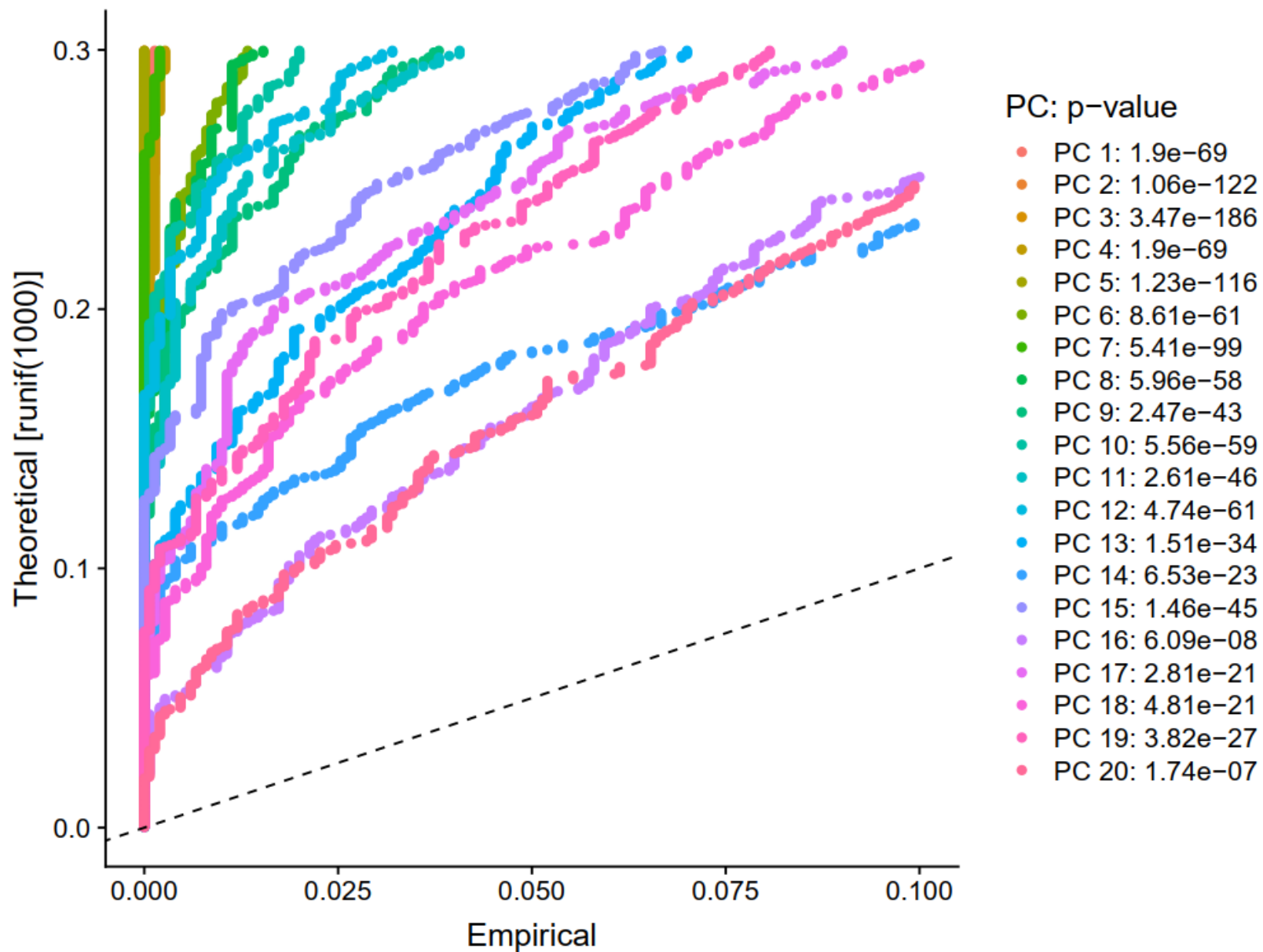
PC\_3



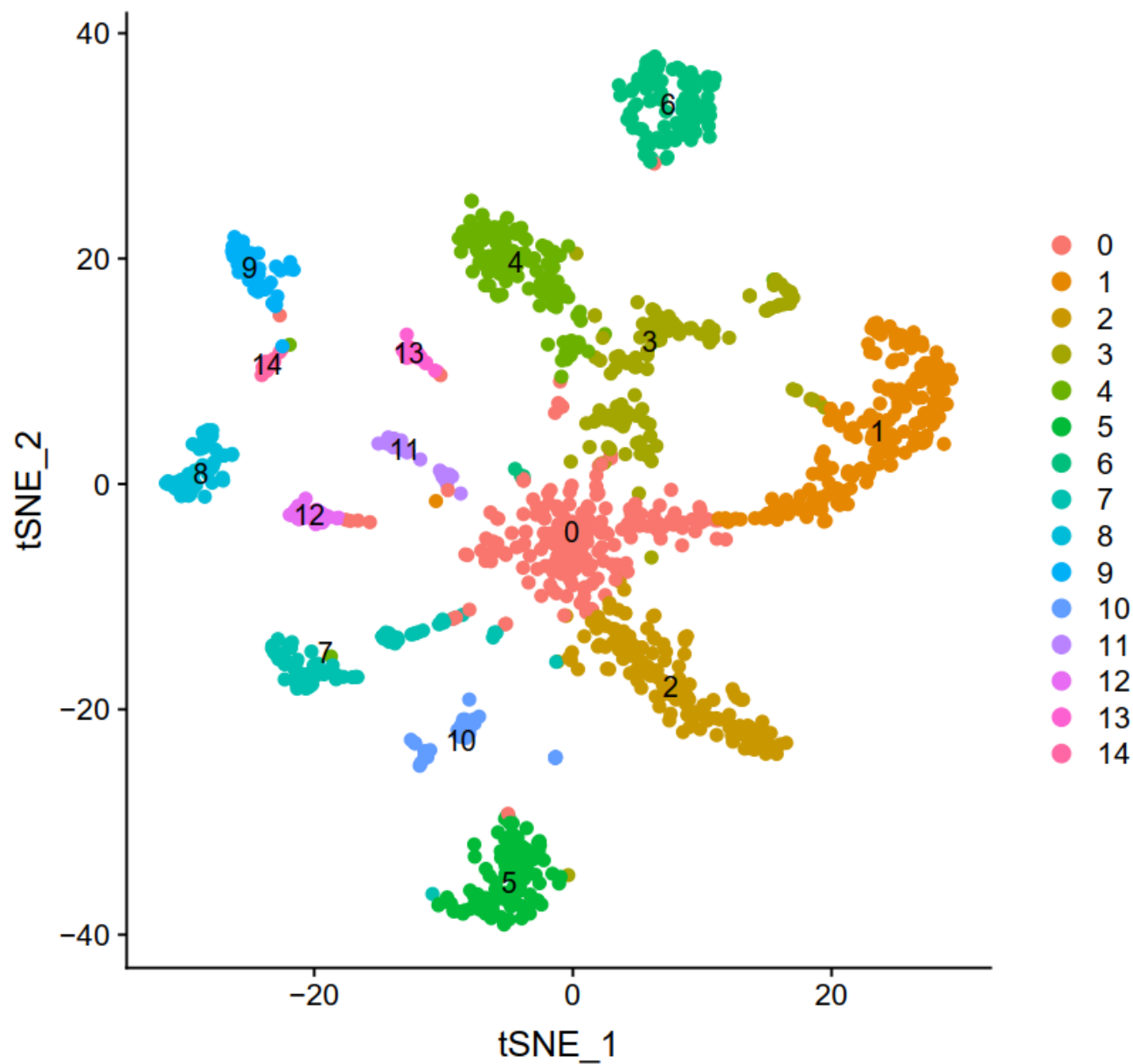
PC\_4



# PCA主成分分析



# TSNE聚类分析



Sample	Cluster
PT039_P11_A01_S1	0
PT039_P11_A02_S2	4
PT039_P11_A03_S3	4
PT039_P11_A04_S4	4
PT039_P11_A05_S5	0
PT039_P11_A06_S6	12
PT039_P11_A09_S9	0
PT039_P11_A10_S10	0
PT039_P11_A11_S11	3
PT039_P11_A12_S12	4

# TSNE聚类分析

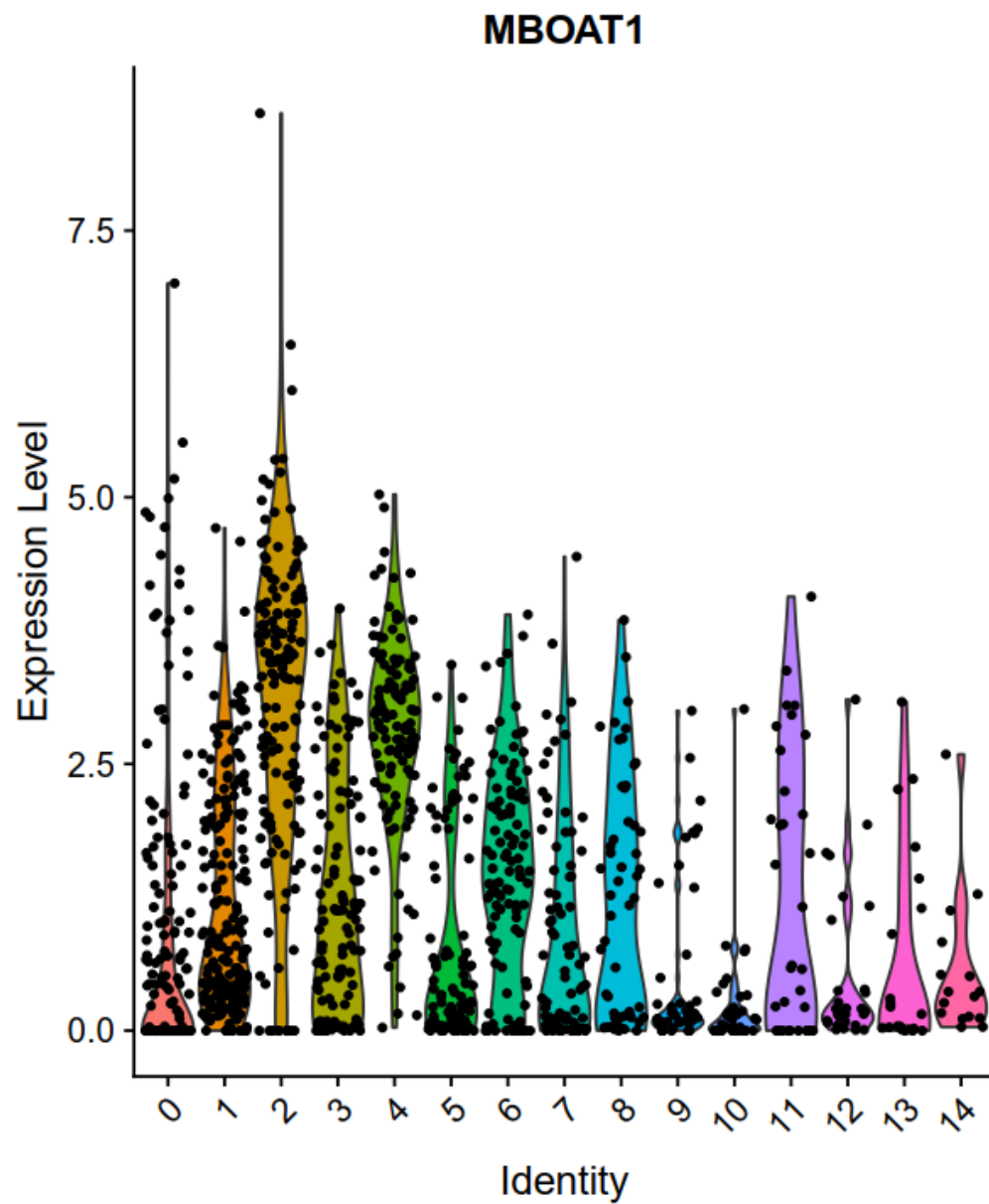
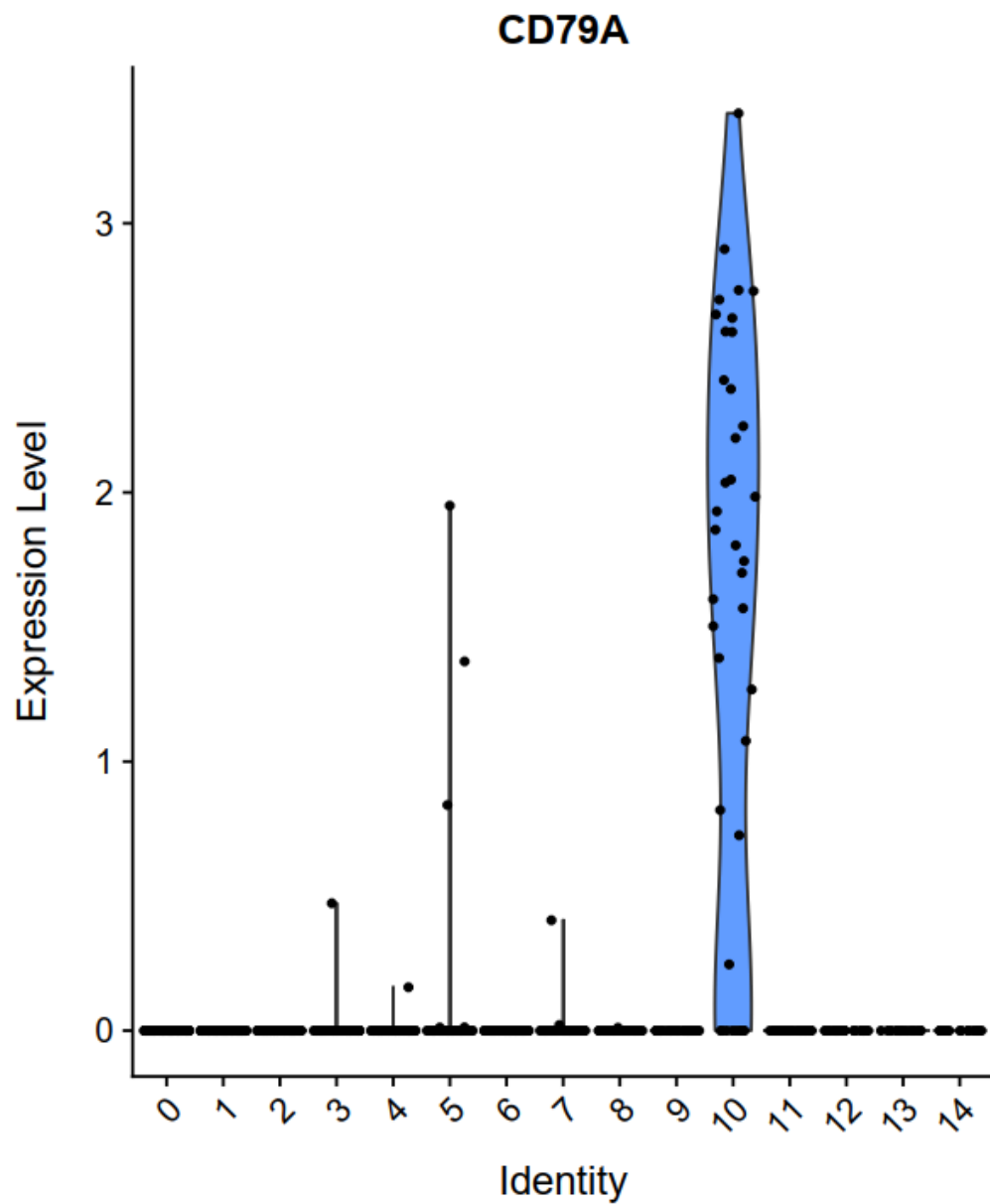




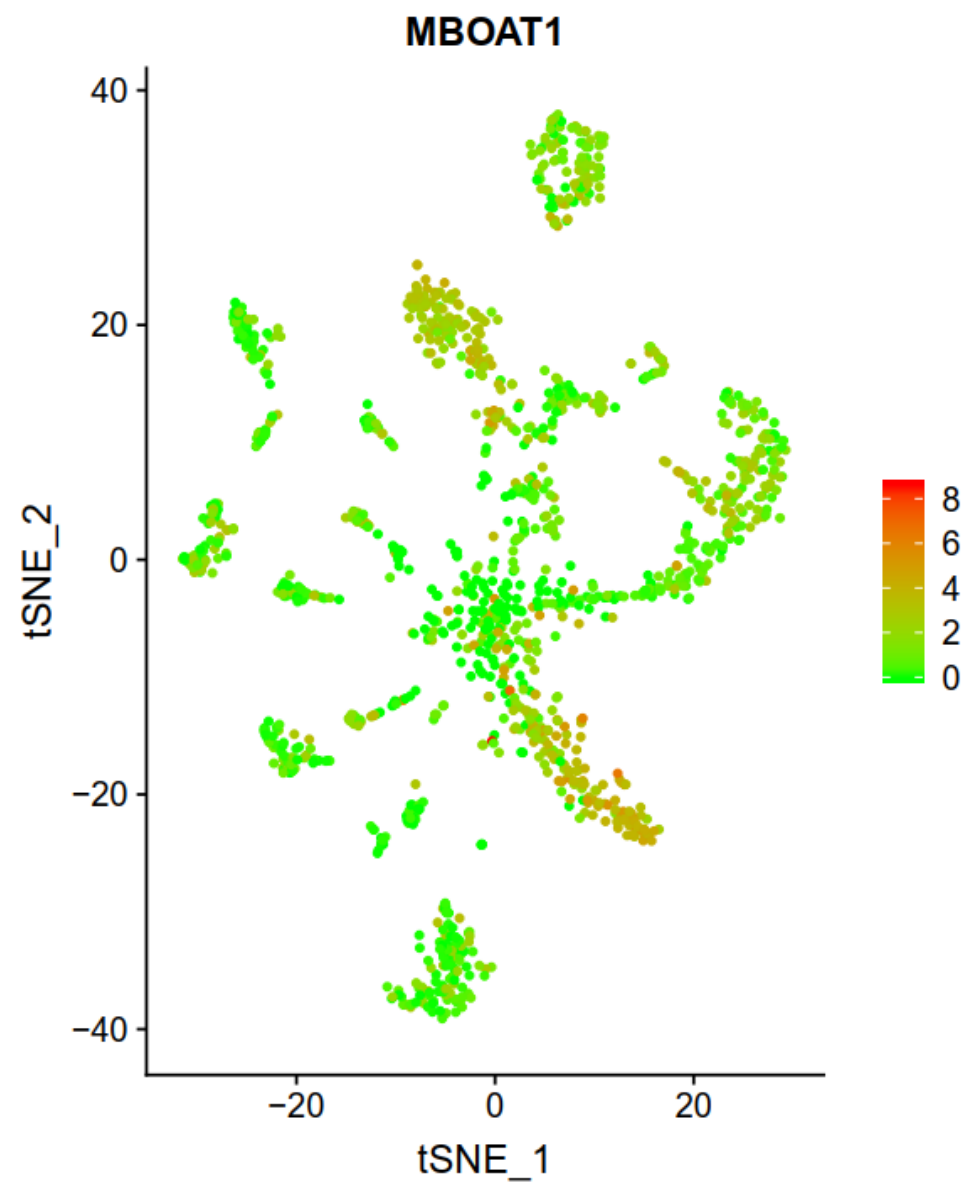
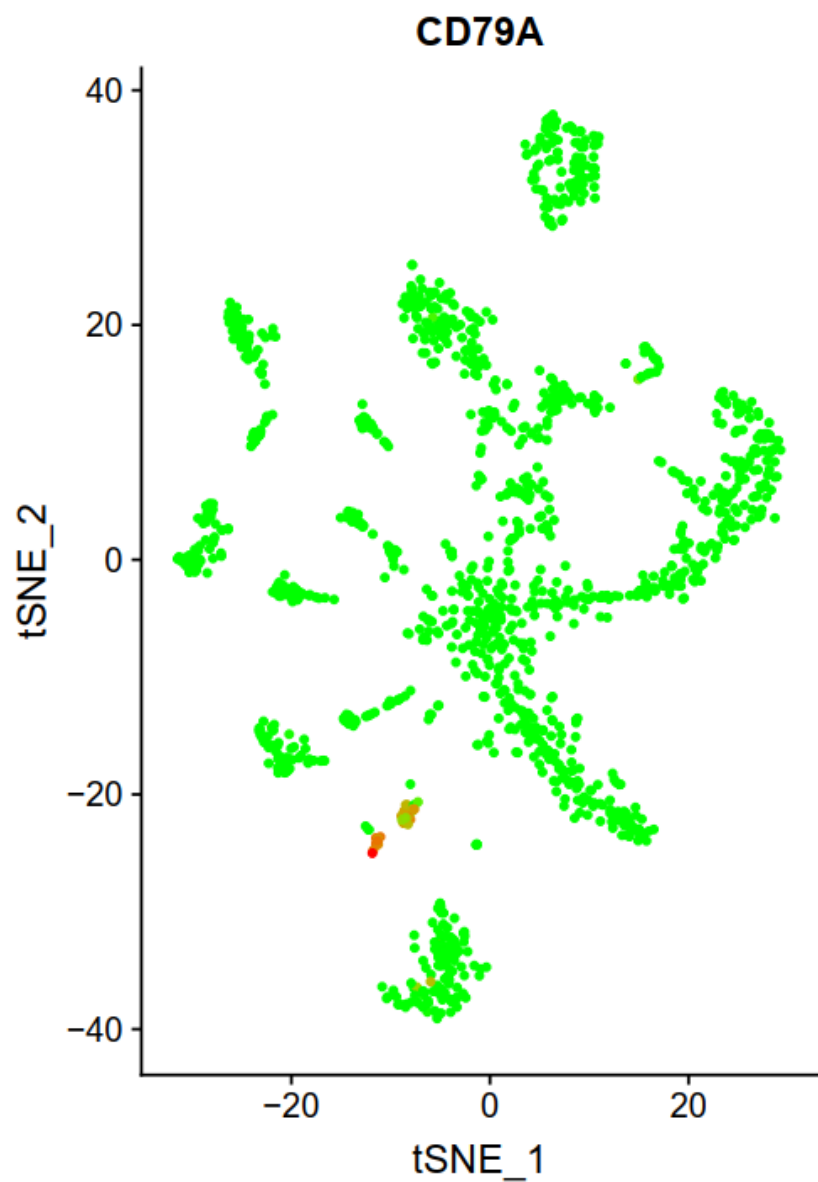
# marker基因

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
2.63E-54	-1.01749	0.21	0.862	4.72E-50	0	GPI
2.20E-50	-1.3175	0.155	0.779	3.95E-46	0	YWHAZ
1.33E-49	-0.99874	0.205	0.879	2.39E-45	0	PNPT1
1.60E-49	-1.39849	0.085	0.708	2.87E-45	0	DDX3X
1.71E-48	-0.56762	0.165	0.816	3.08E-44	0	TMEM41B
1.94E-48	-0.58087	0.165	0.813	3.48E-44	0	S1PR2
2.43E-48	-0.71499	0.135	0.769	4.37E-44	0	ARPP19
3.39E-48	-0.87372	0.135	0.751	6.08E-44	0	MAPK1IP1L
3.53E-48	-0.61756	0.15	0.782	6.34E-44	0	SENP5
3.60E-48	-0.57107	0.145	0.774	6.46E-44	0	CACUL1
4.51E-48	-0.96791	0.14	0.761	8.10E-44	0	PAFAH1B2

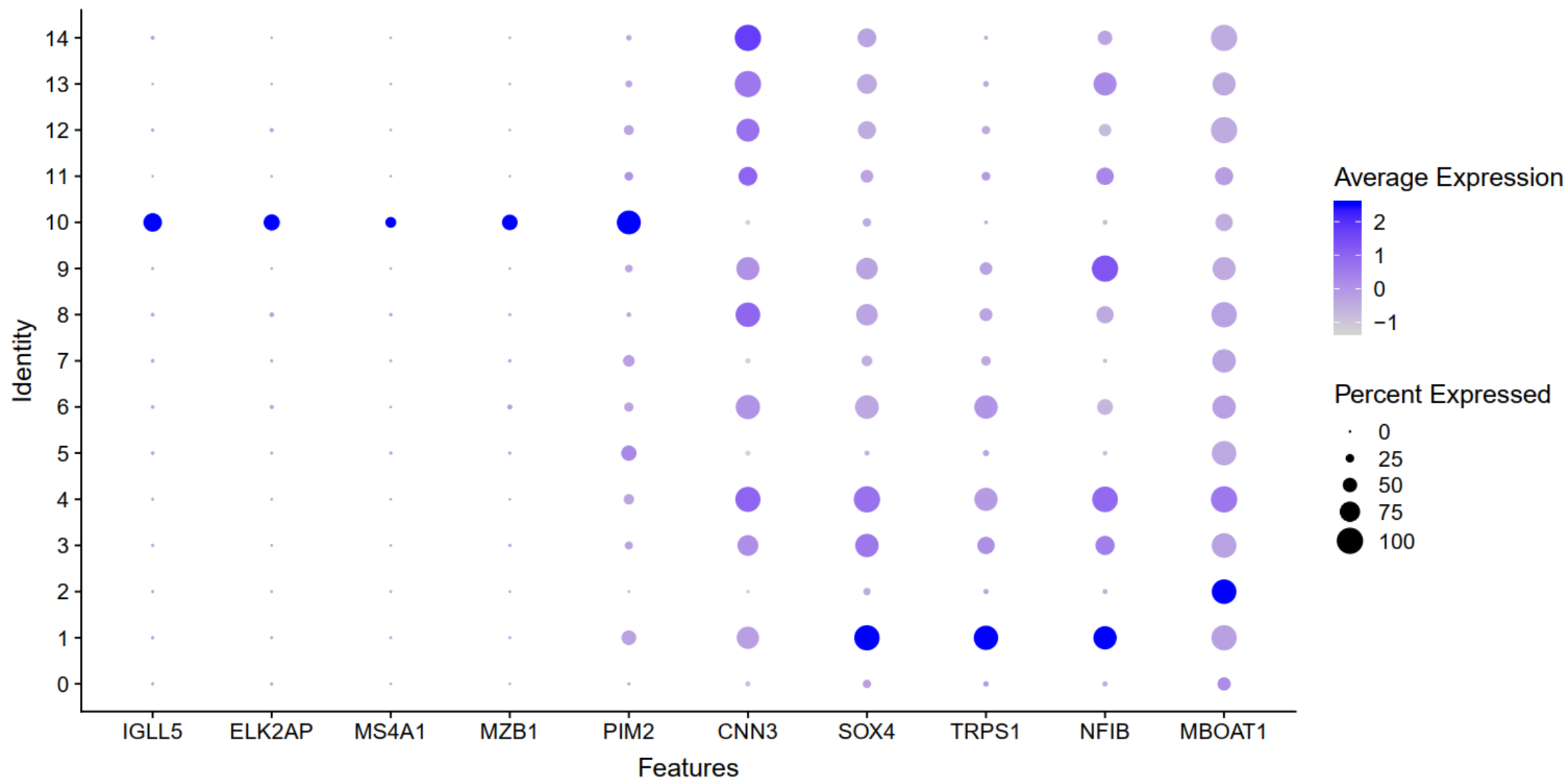
# marker基因



# marker基因



# marker基因

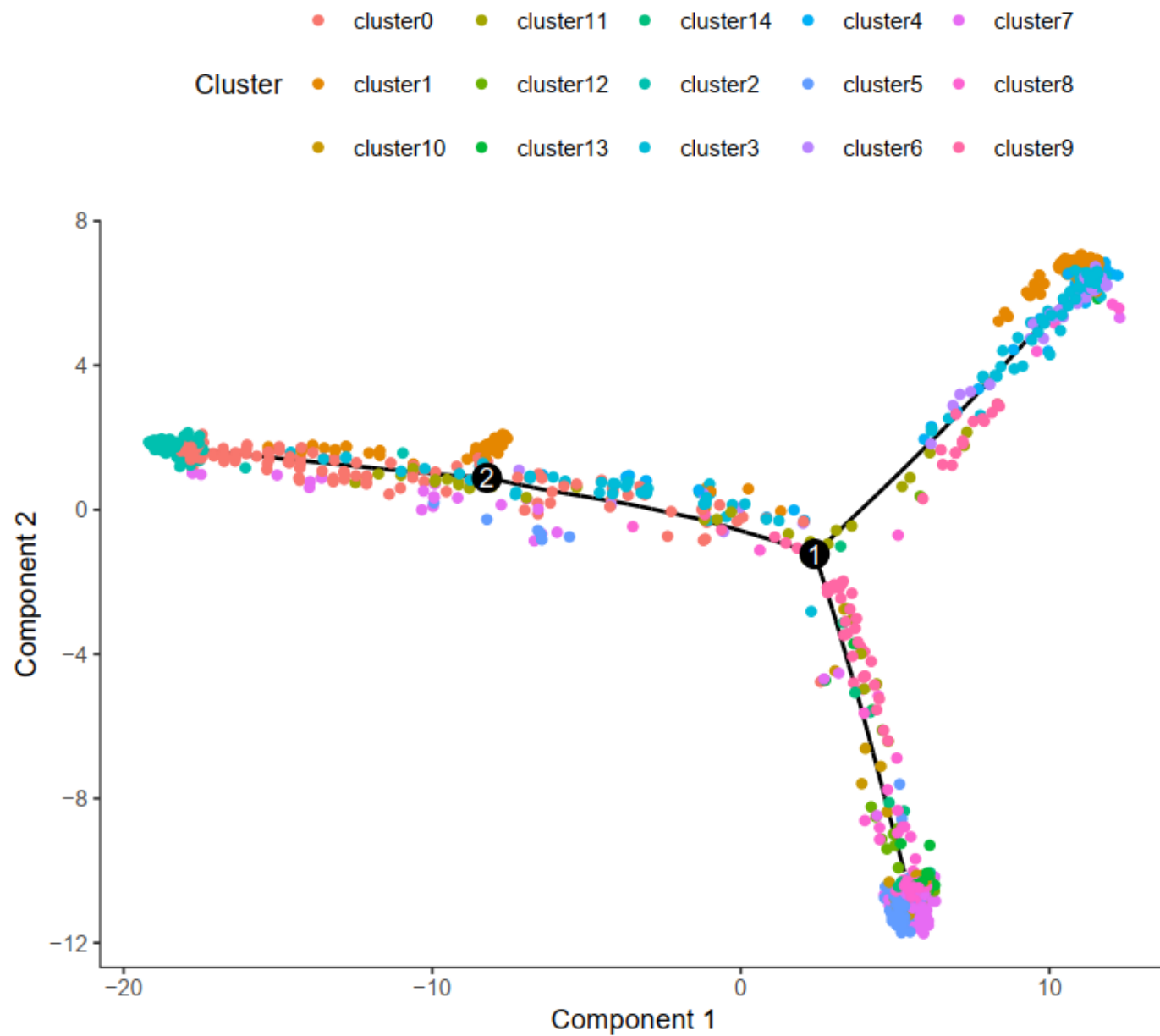


# 细胞类型的注释

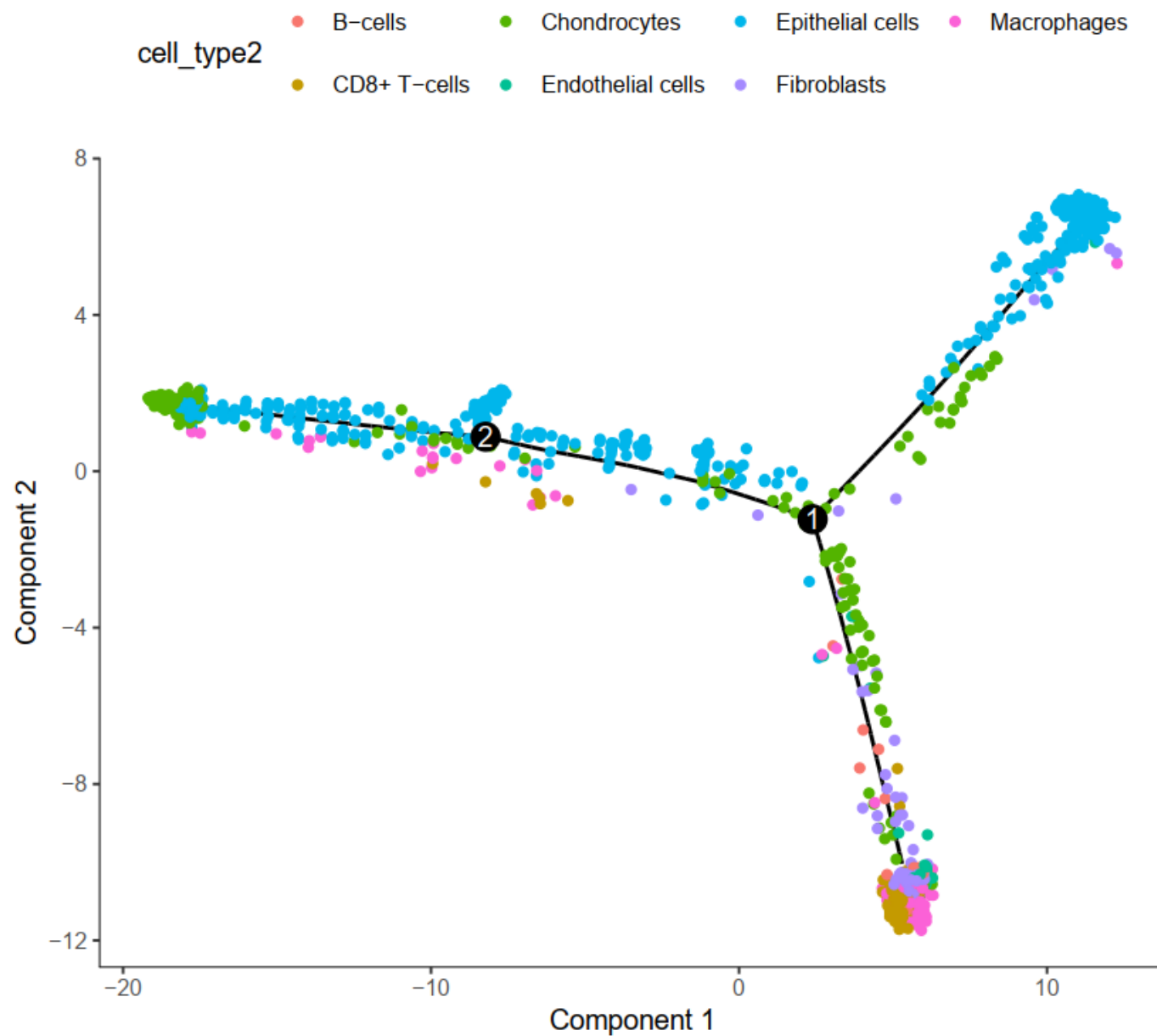
cluster	Type
0	Epithelial cells
1	Epithelial cells
2	Chondrocytes
3	Epithelial cells
4	Epithelial cells
5	CD8+ T-cells
6	Epithelial cells
7	Macrophages
8	Fibroblasts
9	Chondrocytes
10	B-cells
11	Chondrocytes
12	Chondrocytes
13	Endothelial cells
14	Fibroblasts

Cell	Type
PT039_P11_A01_S1	Megakaryocytes
PT039_P11_A02_S2	Megakaryocytes
PT039_P11_A03_S3	Megakaryocytes
PT039_P11_A04_S4	Megakaryocytes
PT039_P11_A05_S5	CD4 T cells
PT039_P11_A06_S6	Megakaryocytes
PT039_P11_A09_S9	CD4 T cells
PT039_P11_A10_S10	CD4 T cells
PT039_P11_A11_S11	Megakaryocytes
PT039_P11_A12_S12	Megakaryocytes
PT039_P11_B01_S13	NA
PT039_P11_B02_S14	Megakaryocytes
PT039_P11_B03_S15	Megakaryocytes
PT039_P11_B04_S16	Megakaryocytes

# 细胞轨迹分析



# 细胞轨迹分析



# 基因名字转换基因id

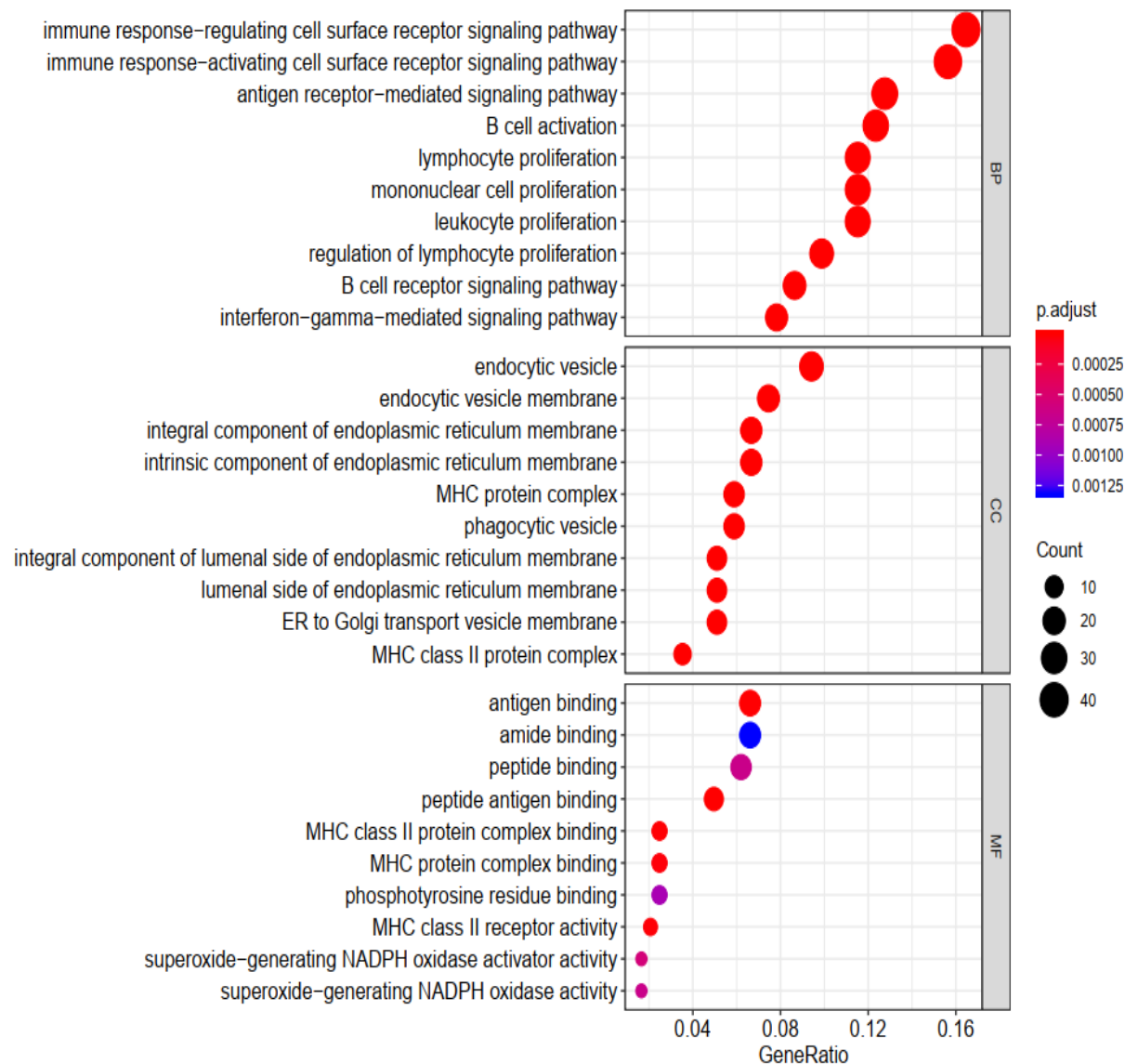
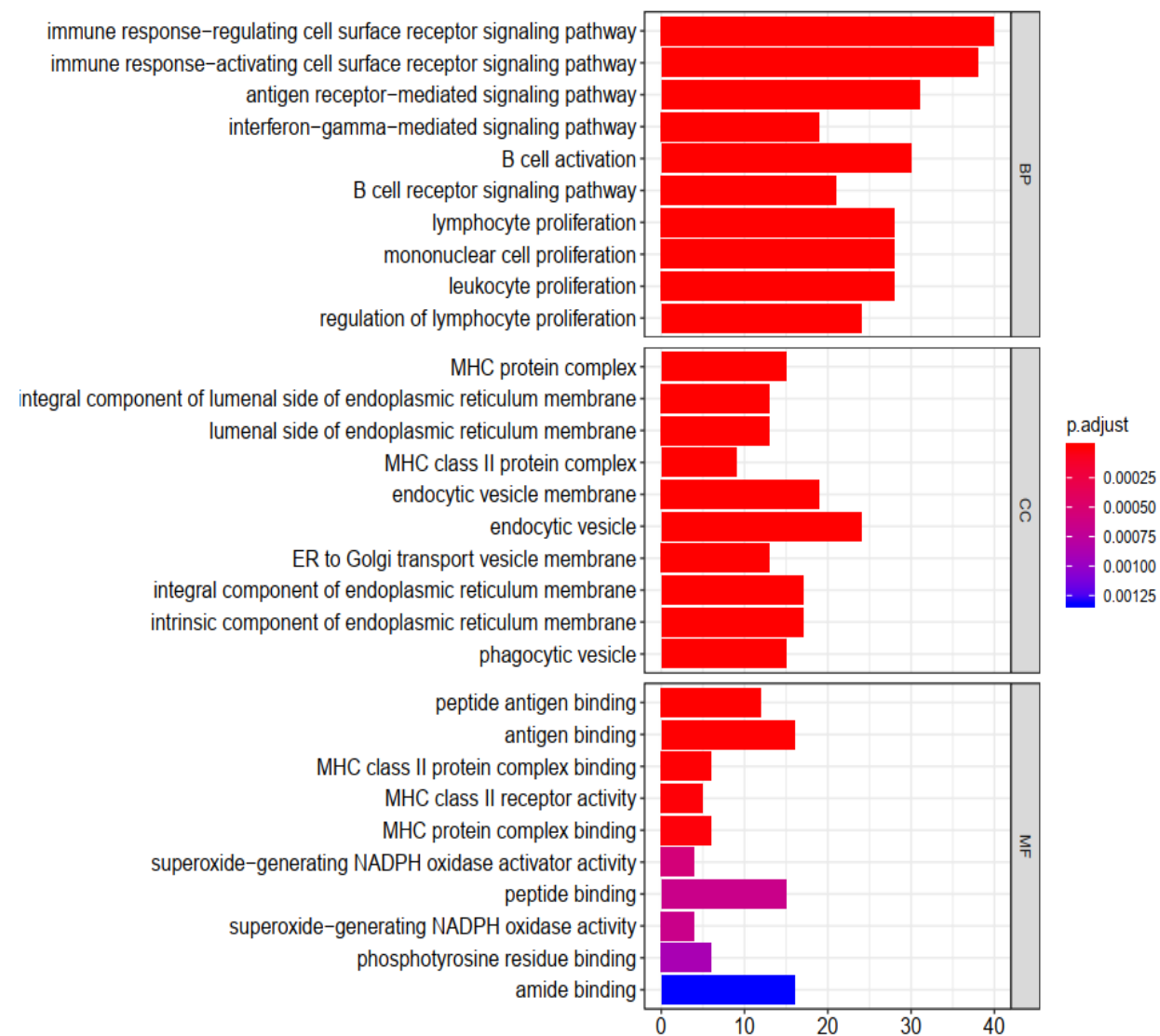
gene	avg_logFC
CD79A	1.952685
FCRLA	1.458243
FCRL5	1.913248
IGLL5	4.629759
CD19	1.391465
P2RX1	0.728547
MZB1	2.562561
TNFRSF17	1.057942
TNFRSF13B	0.931793
ELK2AP	4.089869



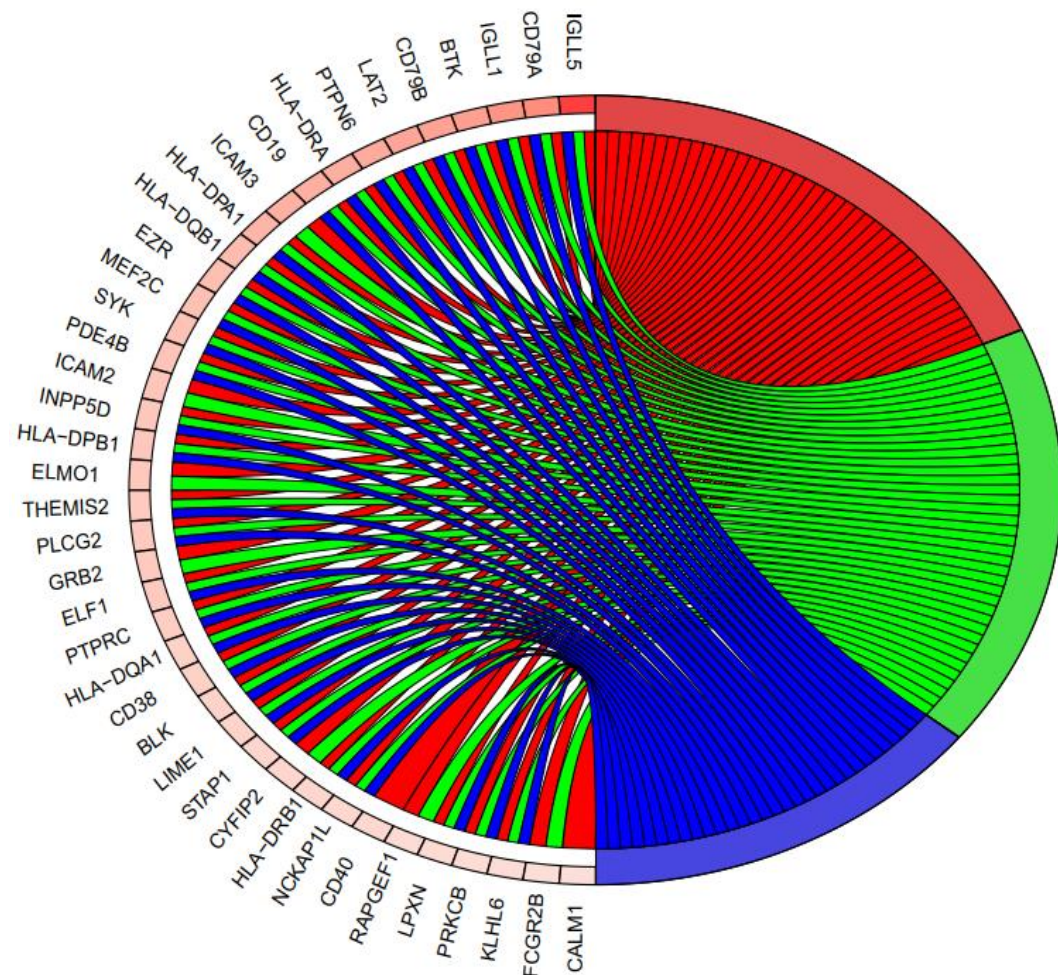
gene	avg_logFC	entrezID
CD79A	1.952685	973
FCRLA	1.458243	84824
FCRL5	1.913248	83416
IGLL5	4.629759	1E+08
CD19	1.391465	930
P2RX1	0.728547	5023
MZB1	2.562561	51237
TNFRSF17	1.057942	608
TNFRSF13B	0.931793	23495
ELK2AP	4.089869	2003



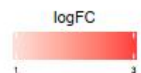
# GO富集分析



# GO卷图



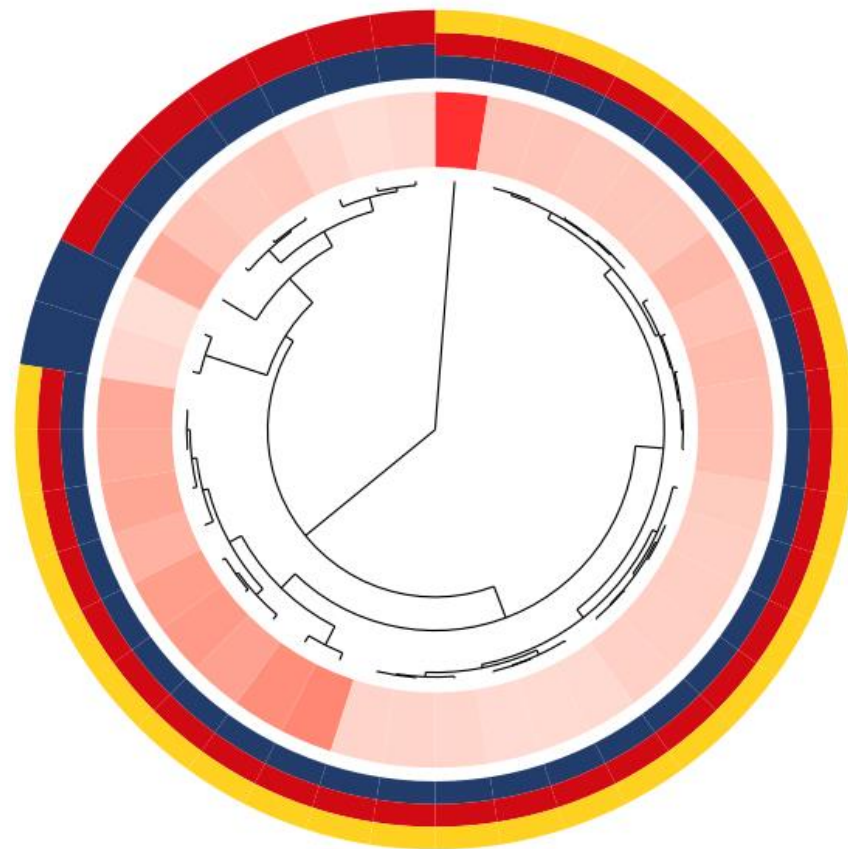
GO Terms ■ immune response-regulating cell surface receptor signaling pathway ■ immune response-activating cell surface receptor signaling pathway ■ antigen receptor-mediated signaling pathway



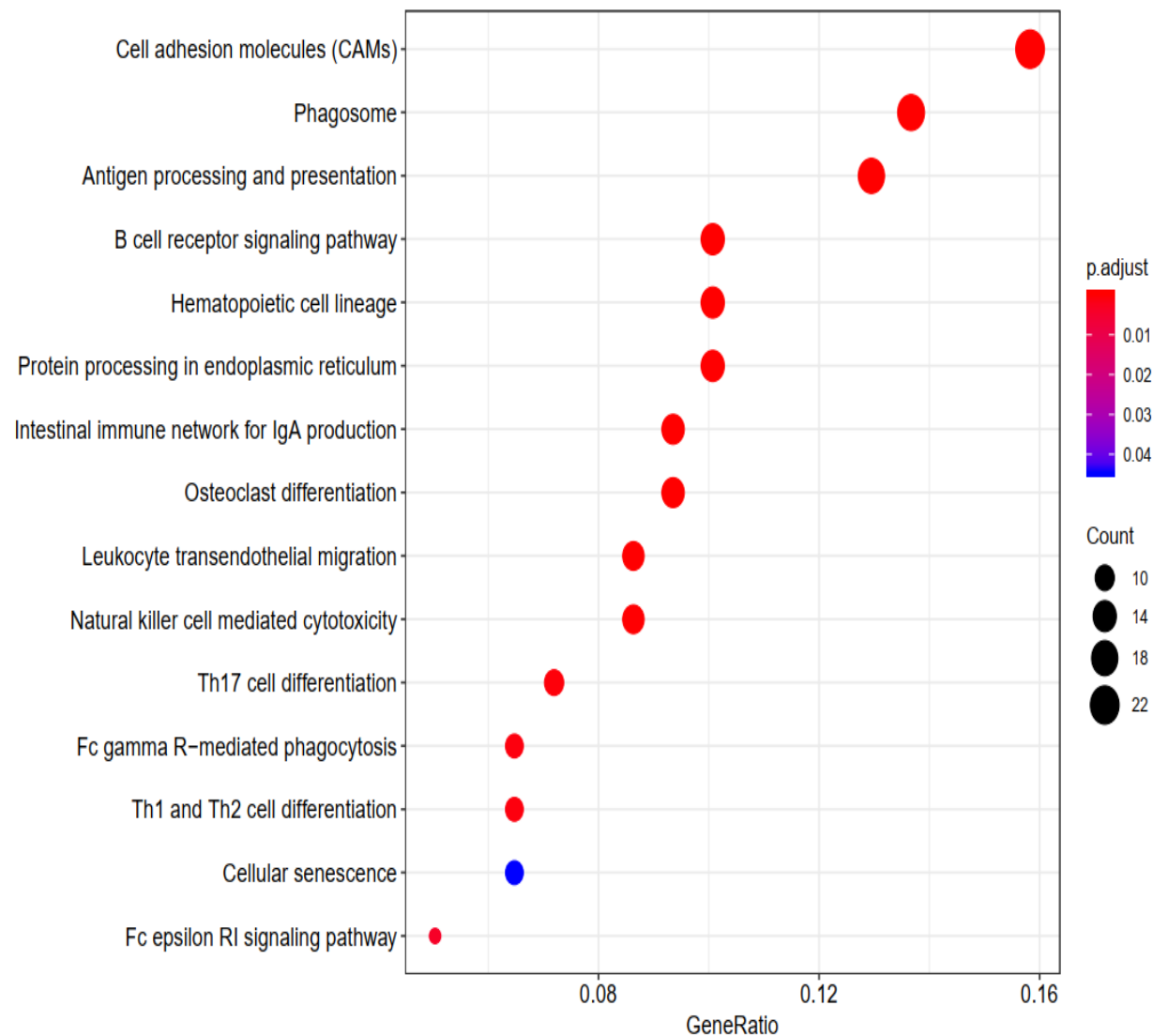
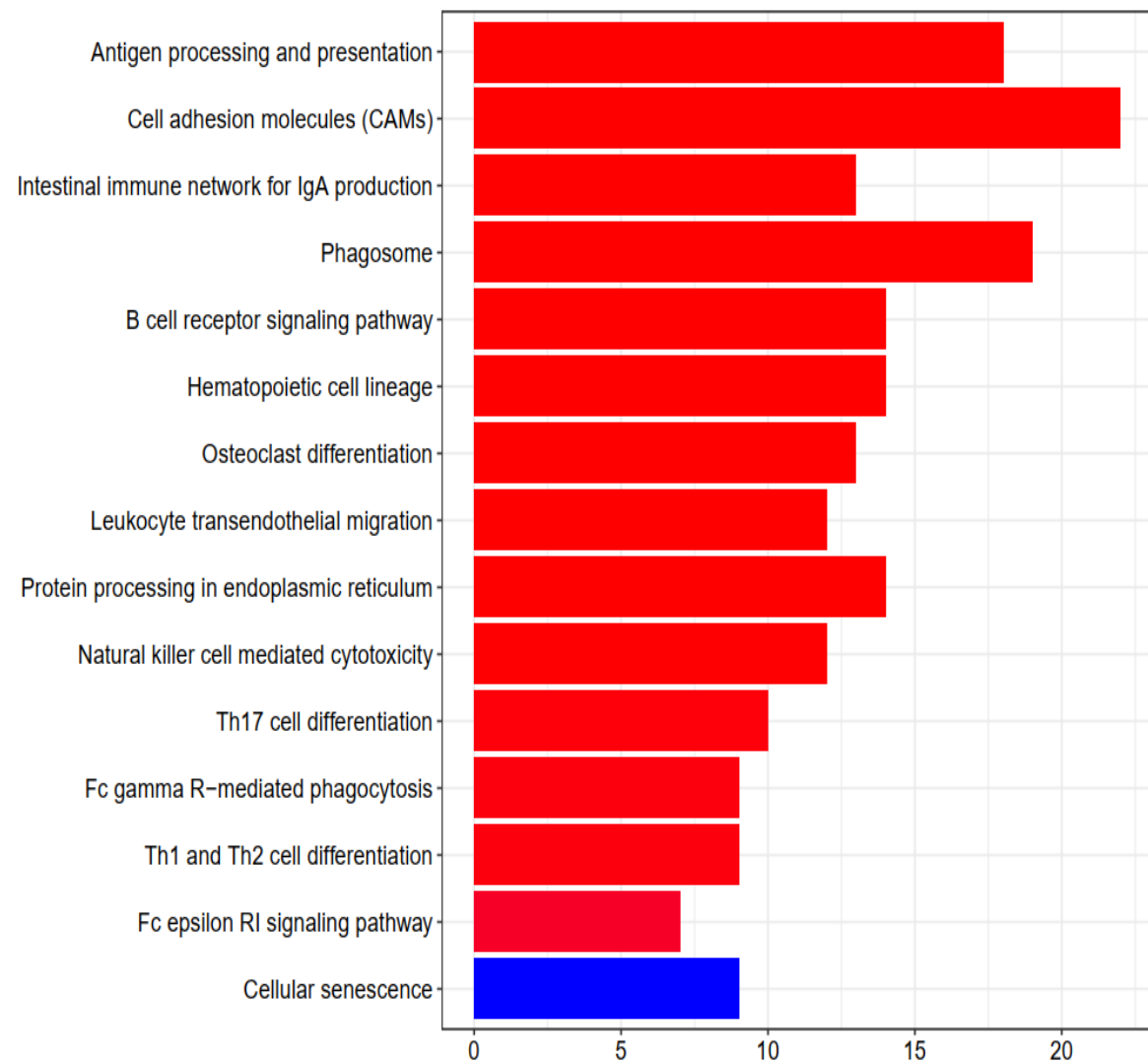
■ immune response-regulating cell surface receptor signaling pathway

■ immune response-activating cell surface receptor signaling pathway

■ antigen receptor-mediated signaling pathway

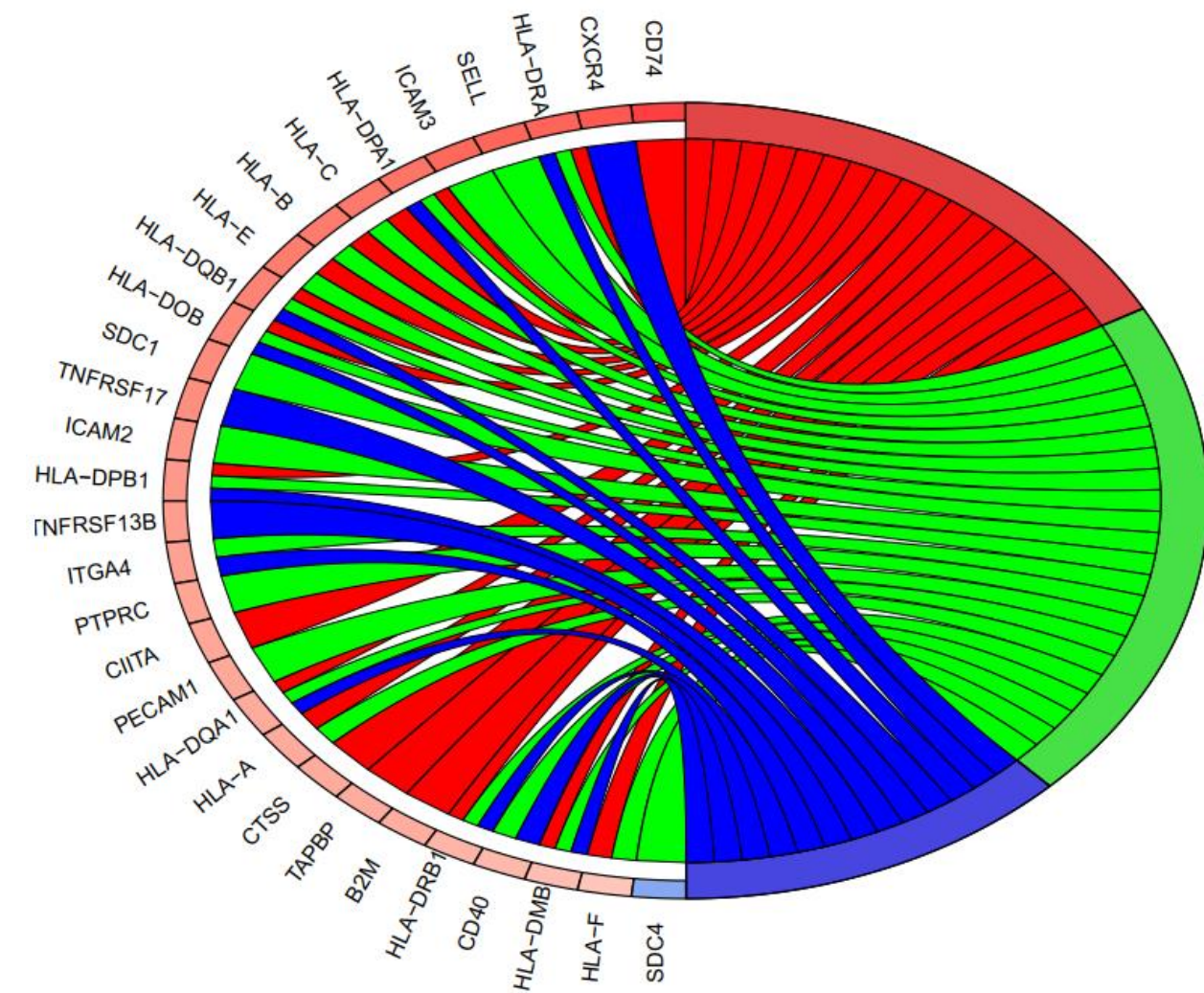


# KEGG富集分析

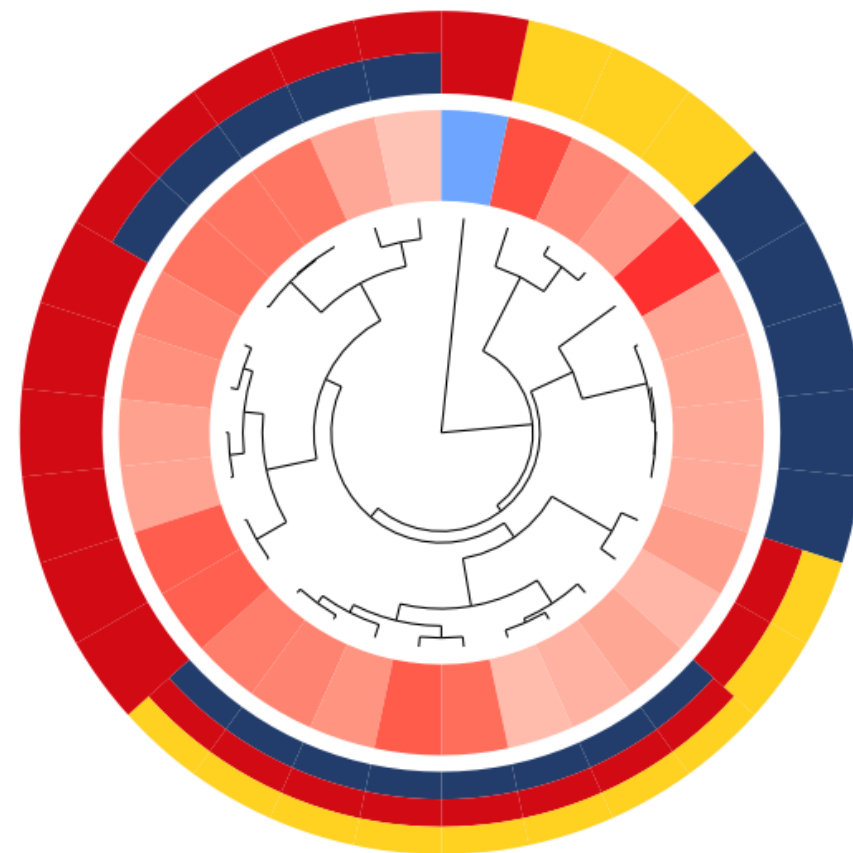




# KEGG 卷图



GO Terms ■ Antigen processing and presentation ■ Cell adhesion molecules (CAMs) ■ Intestinal immune network for IgA production



GO Terms ■ Antigen processing and presentation ■ Cell adhesion molecules (CAMs) ■ Intestinal immune network for IgA production



Thanks!!!