

空间转录组中的机器学习

Machine Learning in Spatial Transcriptomics

www.genechem.com.cn

胡亚辰

2022-5-11

目录 CONTENTS

01. 基于聚类算法的空间分区

02. 基于解卷积算法的细胞类型占比估计

03. 空间配体受体互作分析

04. 空间拟时间分析

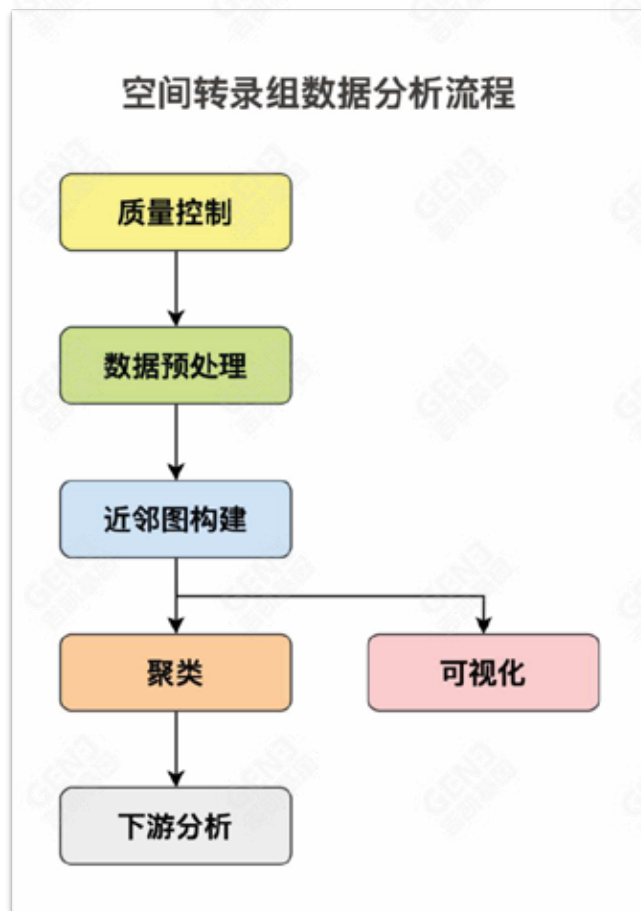
05. 优秀案例文献分享

基于聚类算法的空间分区

Spatial Partition by Using Clustering Method

01.

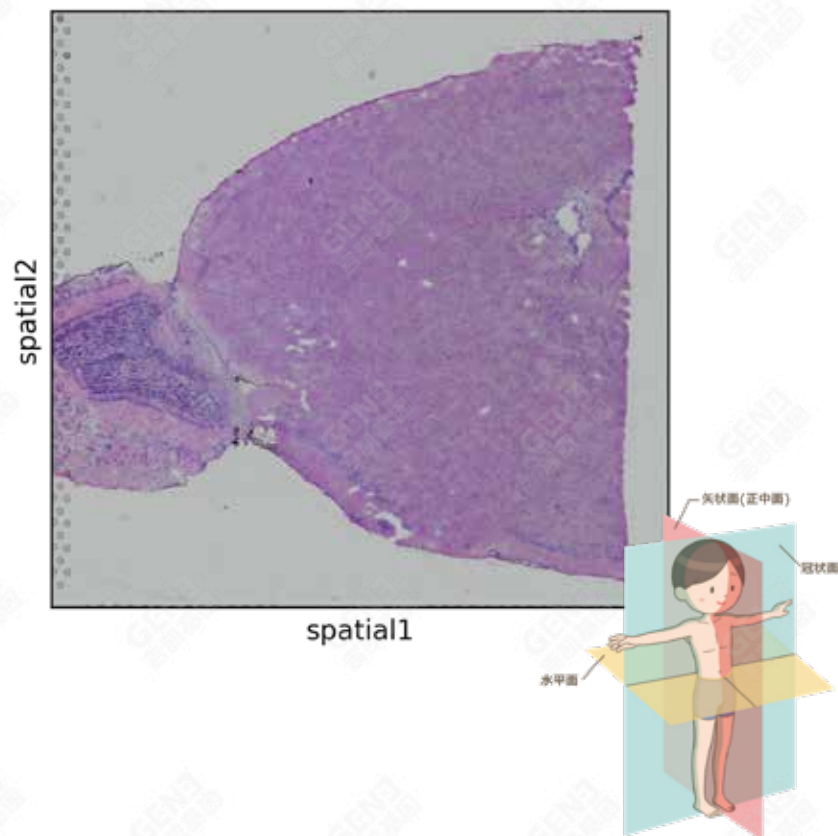
空间转录组数据分析流程



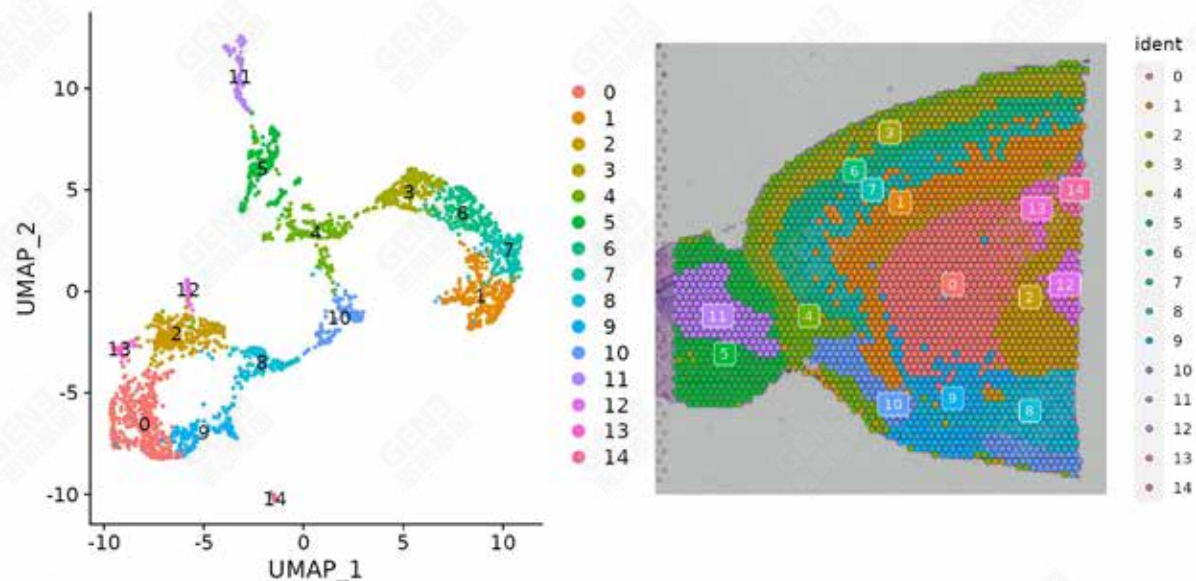
- 空间转录组数据基础分析与单细胞转录组相似
- 分析流程分为质量控制、数据预处理、近邻图构建、聚类和可视化，以及下游分析
- 因为切片已提供物理的空间坐标信息，可用于信息展示。因此无需使用降维算法，如UMAP/tSNE等，进行数据投影和可视化
- 下游分析包括：
 - 细胞类型占比估计（因spot包含多个细胞，不适用细胞类型注释）
 - 空间受体配体互作
 - 空间轨迹推断等

Seurat/Scanpy使用与单细胞转录组一致的流程

小鼠大脑矢状面前部

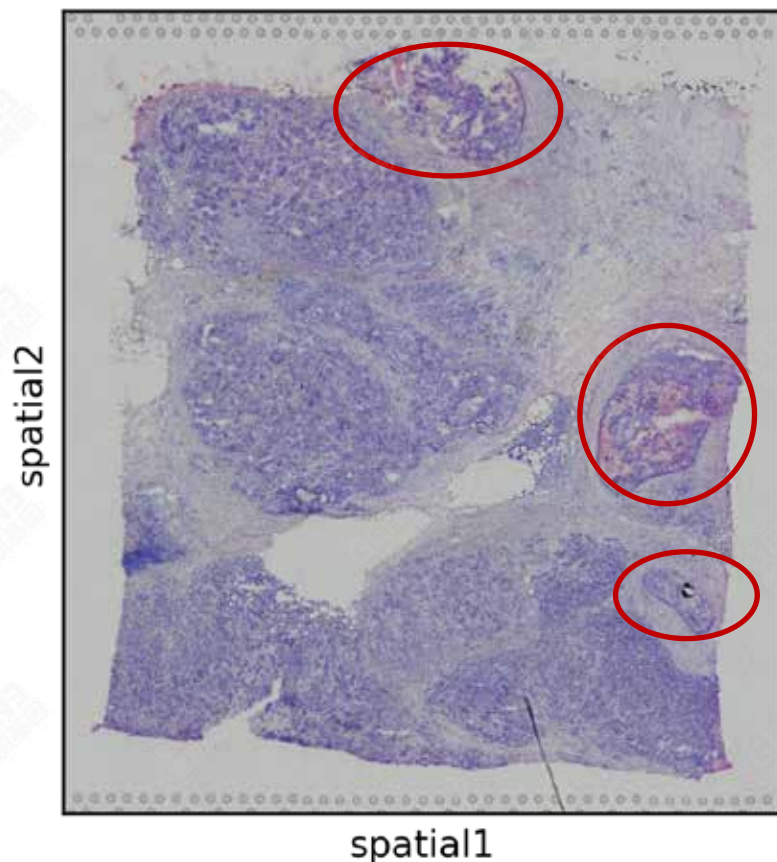


空间转录组数据基础分析与单细胞转录组相似，*Seurat*与*Scanpy*使用与单细胞一致的流程和方法

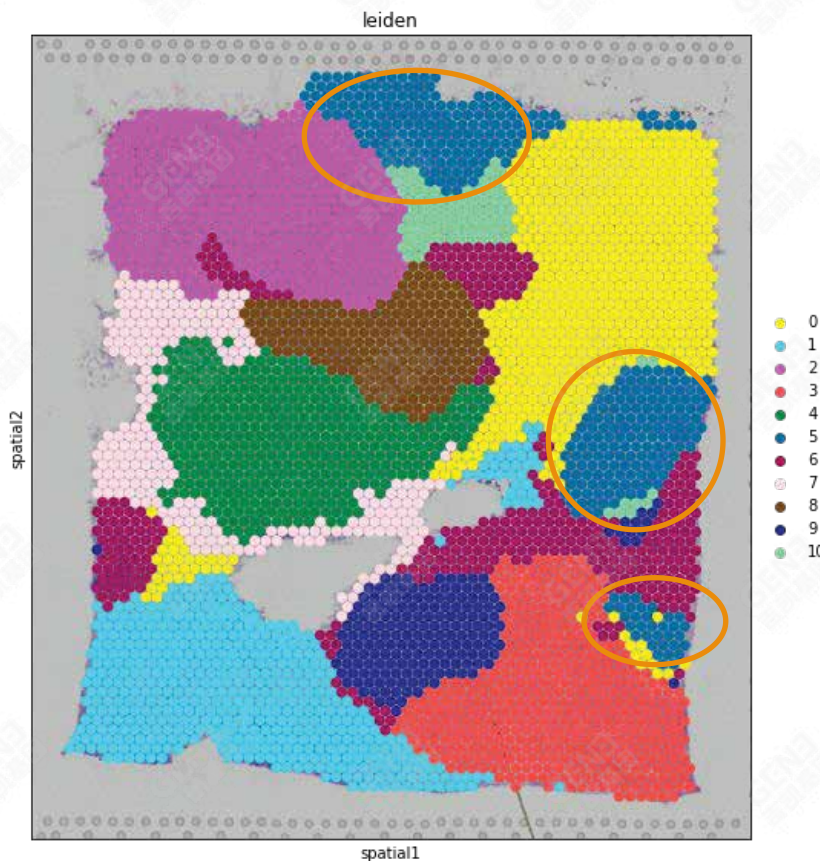


```
brain <- RunPCA(brain, assay = "SCT", verbose = FALSE)
brain <- FindNeighbors(brain, reduction = "pca", dims = 1:30)
brain <- FindClusters(brain, verbose = FALSE)
brain <- RunUMAP(brain, reduction = "pca", dims = 1:30)
```

人乳腺癌



通过形态学观察被判定为DCIS

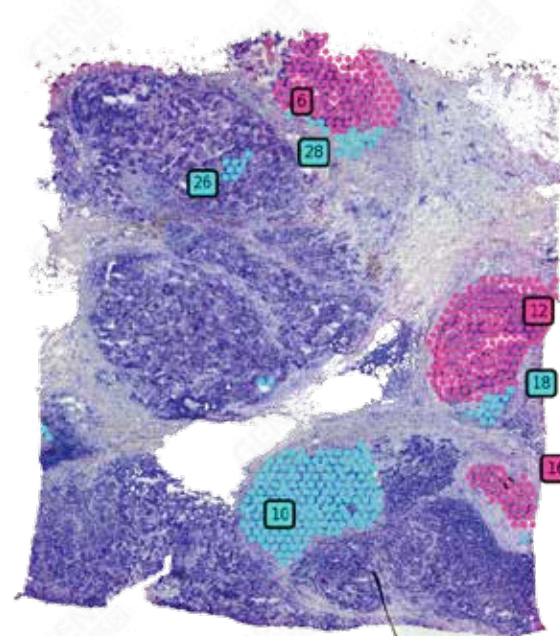
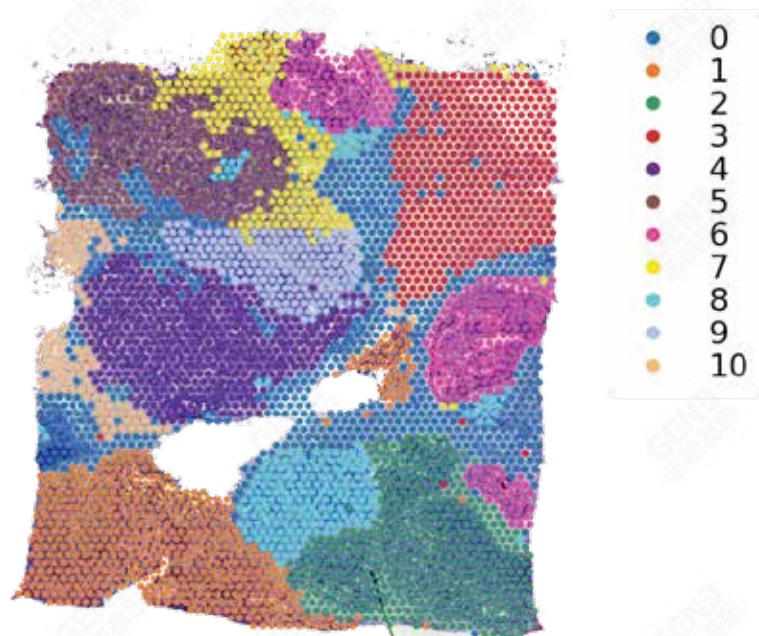


SpaMetric识别为同一cluster的部位

- 一些工具将空间信息纳入考量，如 *SpaMetric*
- 不同于 *Seurat* 使用 **无监督学习** 的方法构建邻近图，*SpaMetric* 使用度量学习 (metric learning) 方法，考虑空间距离信息
- 该方法鉴定出对应于原位导管癌 (DCIS) 的 cluster

```
spm.spatial_reconstruction(adata)  
spm.metric_learning(adata)
```


人乳腺癌



- *stLearn*实现了一种新的方法归一化方法stSME。
- stSME转为空间转录组数据设计的，同事利用了组织空间位置、形态学，以及基因表达信息。
- *StLearn*除了发现原位导管癌（DCIS），还发现其附近的浸润性导管癌（IDC）

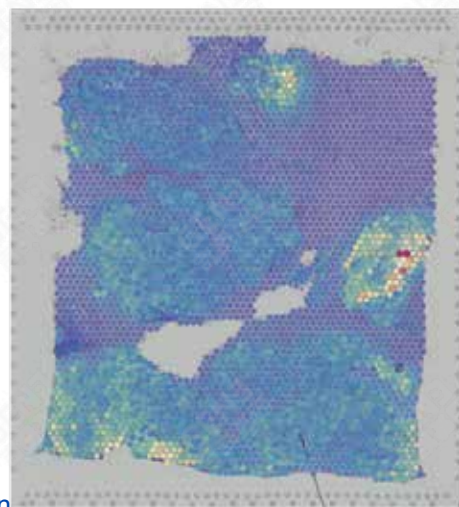
```
# apply stSME to normalise log
transformed data
st.spatial.SME.SME_normalize(data_SME,
use_data="raw")
data_SME.X =
data_SME.obsm['raw_SME_normalized']
st.pp.scale(data_SME)
st.em.run_pca(data_SME,n_comps=50)
```

差异表达分析与Marker基因可视化

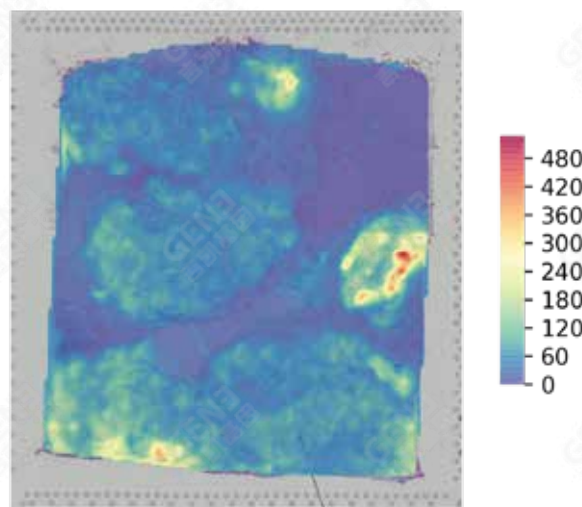
	0	1	2	3	4	5	6	7	8	9	10
0	ACKR1	CXCL14	COX6C	CRISP3	CPB1	MGP	IGKC	C1QA	LINC00645	IFT122	ALB
1	CCL21	CCND1	SNCG	SLITRK6	HLA-B	S100G	IGLC2	APOE	MUC5B	S100G	MGP
2	TNXB	DEGS1	SLC39A6	S100A13	COX6C	DSP	IGHG1	C1QB	PVALB	AC087379.2	BGN
3	CLDN5	S100A11	WFDC2	IGFBP5	IL6ST	TFF1	IGHG3	FABP4	IGFBP2	IFI6	IGHG1
4	CCL19	CPNE7	CSTA	SERHL2	HLA-C	TFF3	C3	GPX3	EXOC2	CALML5	IGFBP7
5	ADAM33	DEGS2	MT-ND1	CALML5	HLA-A	MT-ND3	IGHG4	ADH1B	SLC30A8	HLA-B	COL1A1
6	C7	MUC1	MCCD1	S100A16	ADIRF	RERG	IGLC3	ADIPOQ	COLEC12	HEBP1	SFRP2
7	CHRD1	RPLP1	RAB11FIP1	KRT8	CFB	STC2	IGHA1	FOLR2	ZNF703	TPD52	CST1
8	ADH1B	PSMB4	MT-CO1	S100A11	B2M	MT-CO1	IGHM	COL14A1	SLC39A6	HLA-A	HTRA3
9	SOD3	RPSA	UGCG	NUPR1	H2AFJ	HK2	TIMP1	HSPB6	AC037198.2	ISG15	COL1A2

- 与单细胞转录组分析方法类似，可对空间分区cluster进行差异表达分析，并对得到的marker基因表达水平进行可视化
- 热图和等高线图是常见的表现形式

热图
GAPDH



等高线图
GAPDH



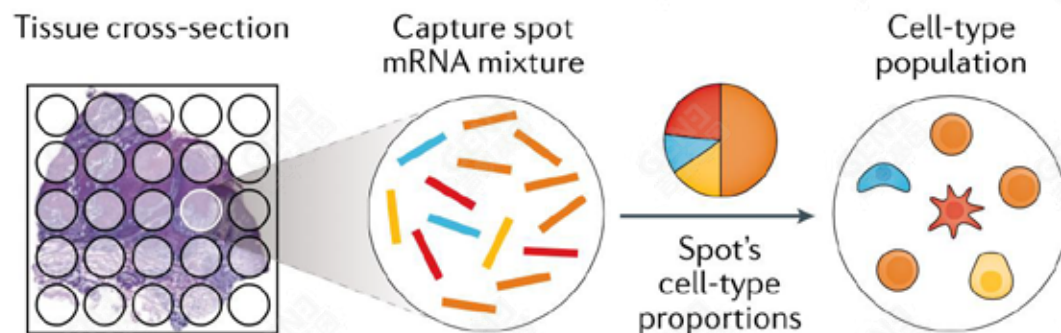
基于解卷积算法的细胞类型占比估计

Cell Type Proportion Estimation Based on Convolution

02.

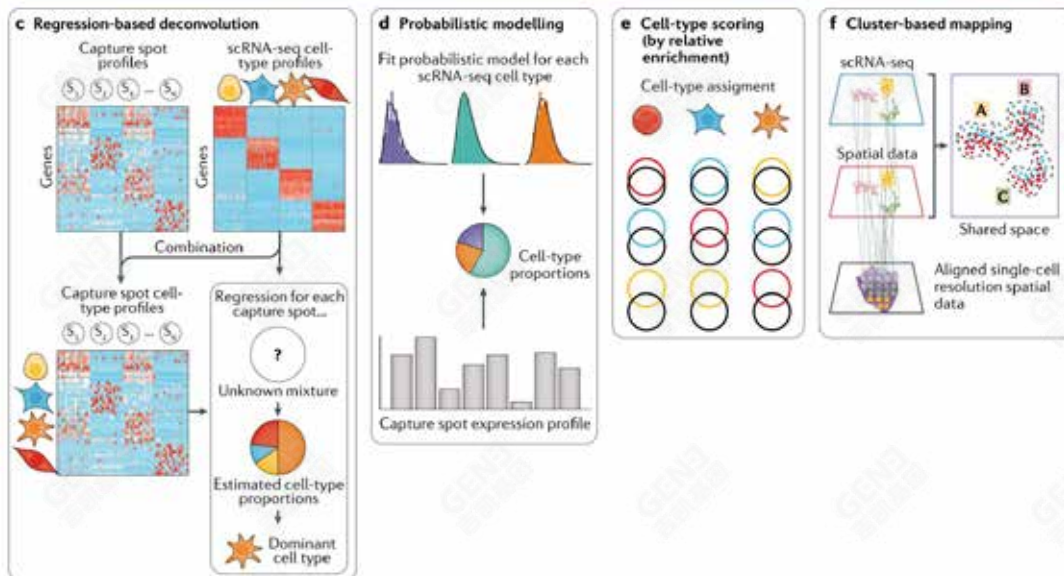
空间转录组为什么要整合单细胞转录组数据分析？

a Deconvolution of spatial barcoding capture spots



Deconvolution

Mapping

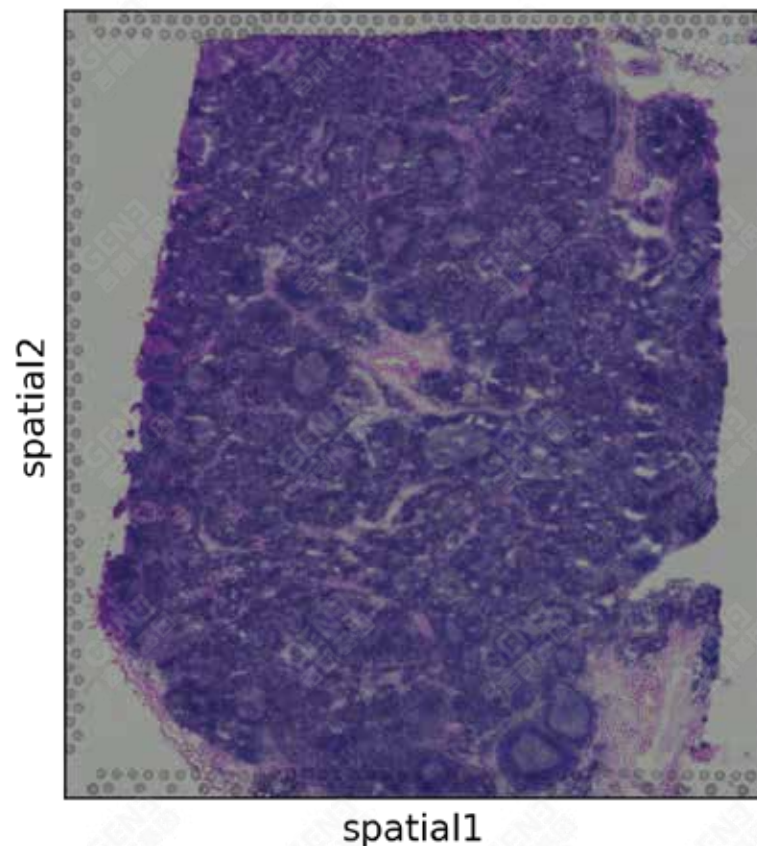


- 10x visium spot约10个细胞。分辨率转录组数据无法区分多个细胞类型。一个分辨多个细胞类型的转录组数据
- 常用方法的原理可分为解卷积与映射两大类。解卷积试图从基于单细胞数据的每个spot的mRNA混合物中分离出离散的细胞亚群
- 为了确保细胞类型组成的一致性，推荐使用同一样本，分别进行空间转录组与单细胞转录组的实验，并对数据整合分析

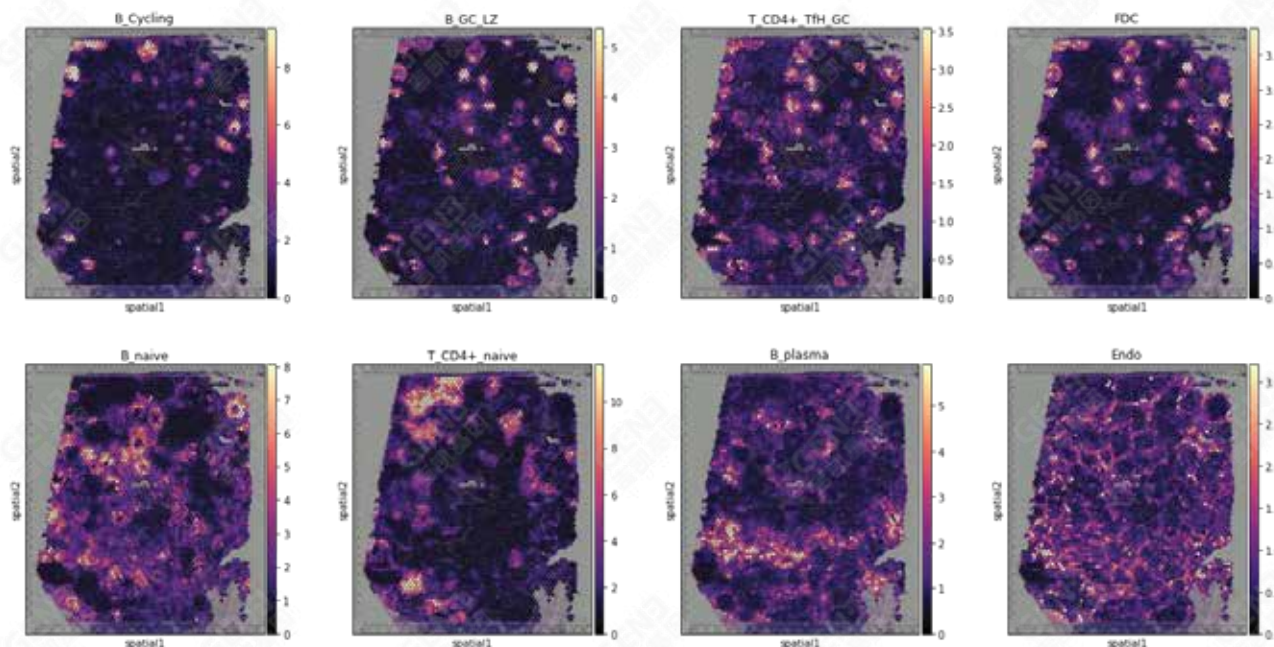


Cell2location使用负二项分布回归

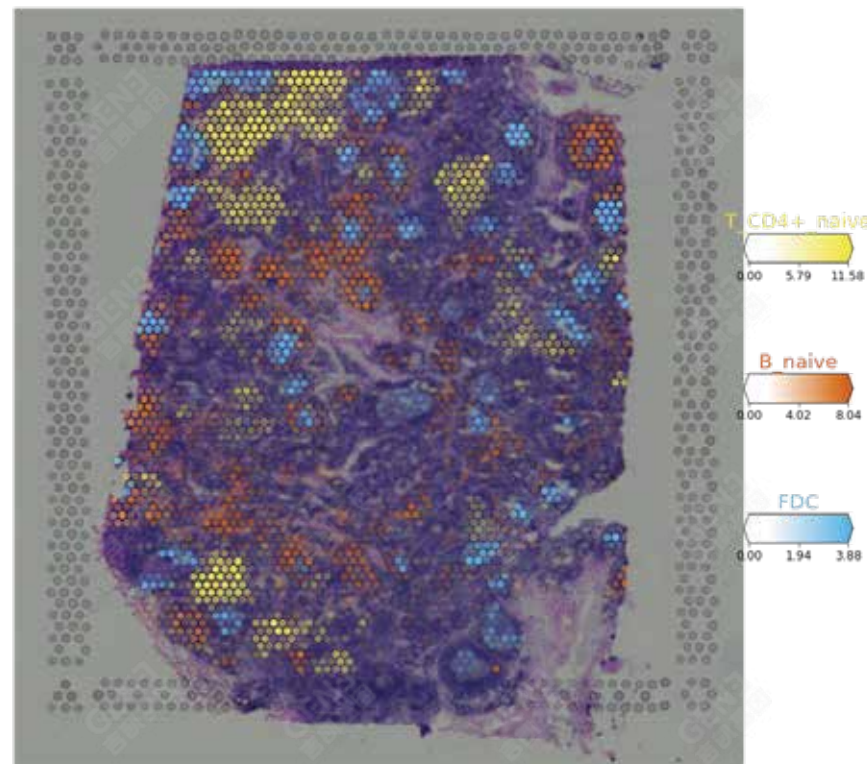
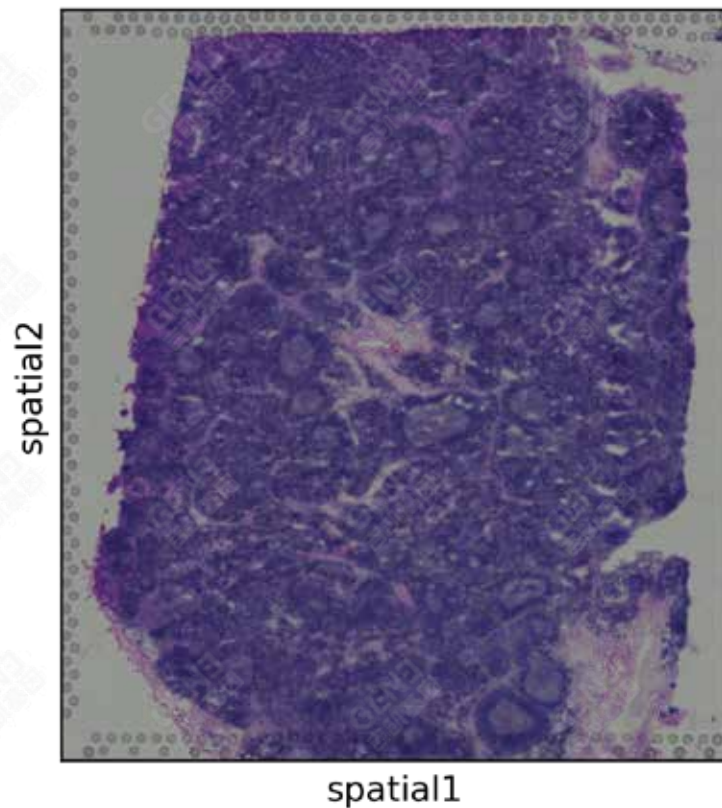
人淋巴结



- Cell2location使用带有细胞类型注释标签的单细胞转录组数据训练负二项分布模型，在将细胞类型信息隐射在空间转录组的spot上
- 人淋巴结切片中，可见不同免疫细胞类型的淋巴小结，以及富含内皮细胞的淋巴管结构



多个细胞类型占比可视化



空间配体受体互作分析

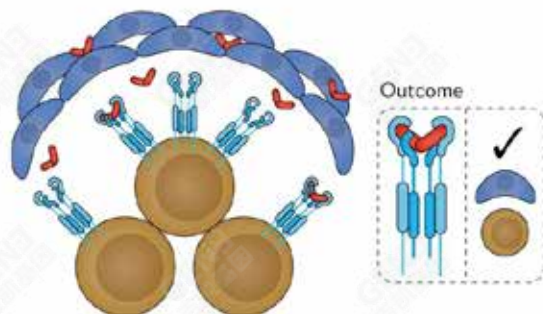
Spatial Ligand-receptor Analysis

03.

不仅考虑LR配对信息，还要考虑物理空间位置

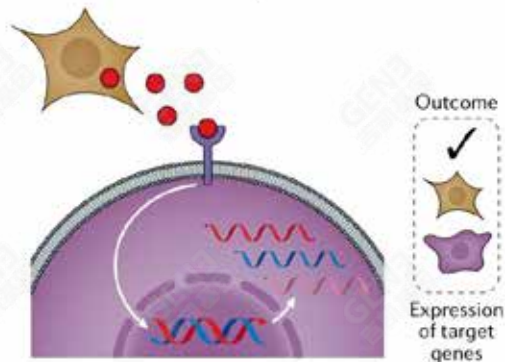
A Restricted by ligand-receptor proximity

Aa Communicative cells

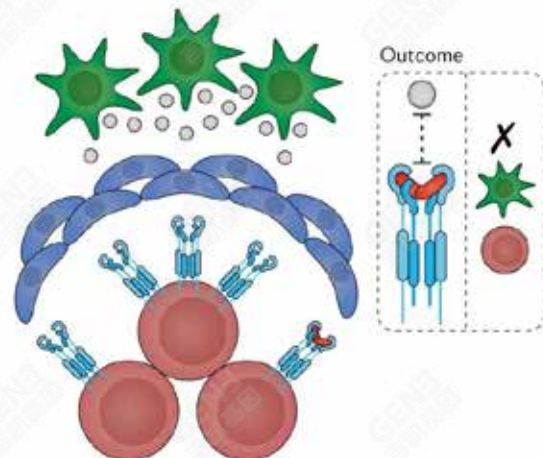


B Restricted by ligand-receptor-target co-expression

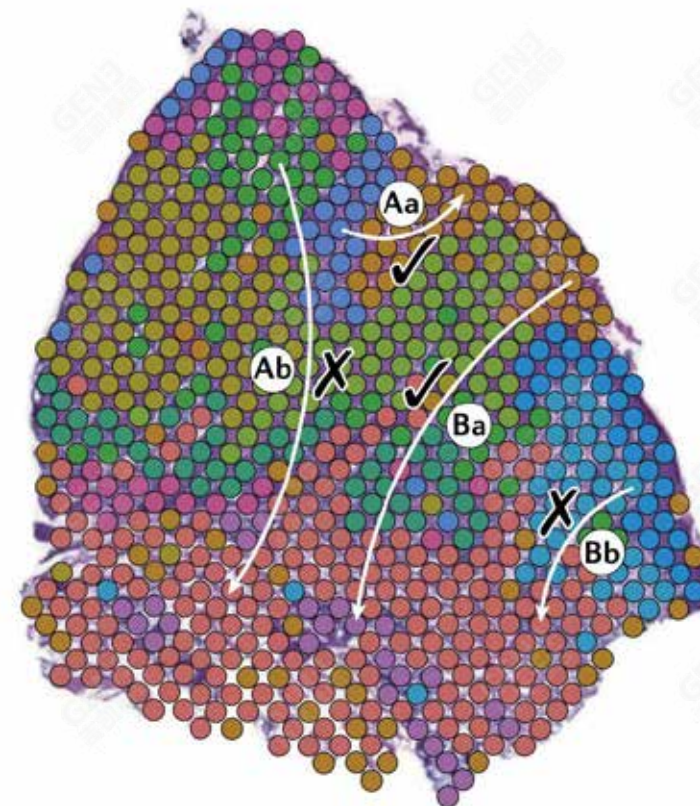
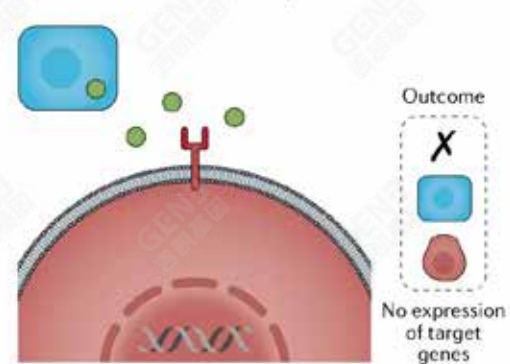
Ba Communicative cell pair



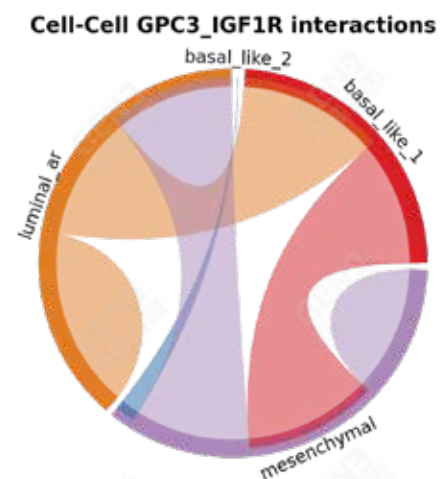
Ab Non-communicative cells



Bb Non-communicative cell pair



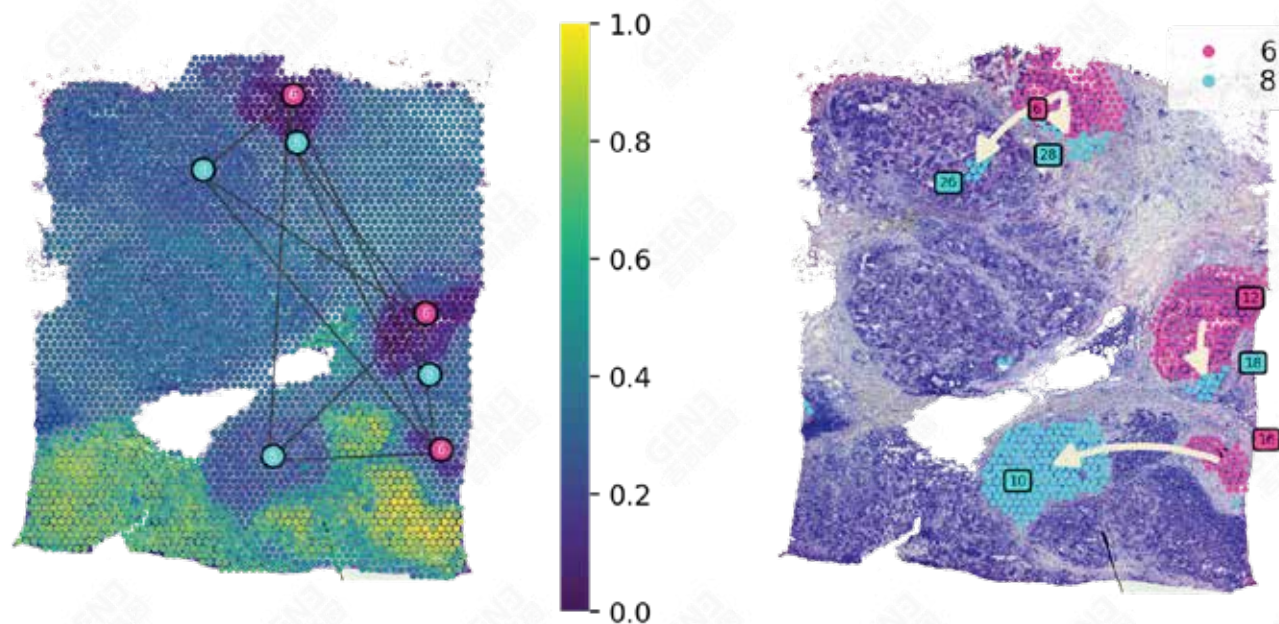
-
- Scatter plot showing the relationship between LR Rank (n sig spots) on the x-axis and n_spots_sig on the y-axis. The plot shows a strong negative correlation, with data points labeled by gene names. The top-left points are labeled with gene names like NCAM, NECTIN4, and others, while the bottom-right points are labeled with gene names like NCAM, NECTIN4, and others. The x-axis ranges from 0 to 50, and the y-axis ranges from 20 to 70.



空间拟时间分析

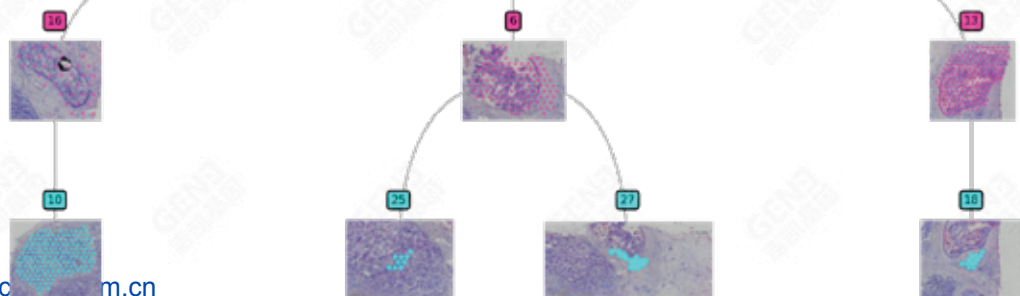
Spatial Pseudotime Analysis

04.



- 单细胞转录组分析的关键点在于计算出能代表细胞分化的空间，并将细胞嵌入（embed）其中。而空间转录组天然地携带物理空间信息
- 与空间转录组拟时序分析类似，需要指定细胞分化的起始位点，也就是根spot
- 在乳腺癌的示例中，拟时间信息与空间位置信息均支持癌症从原位向侵袭性的转化

Pseudoroot



优秀案例文献分享

Collection of Spatial Articles

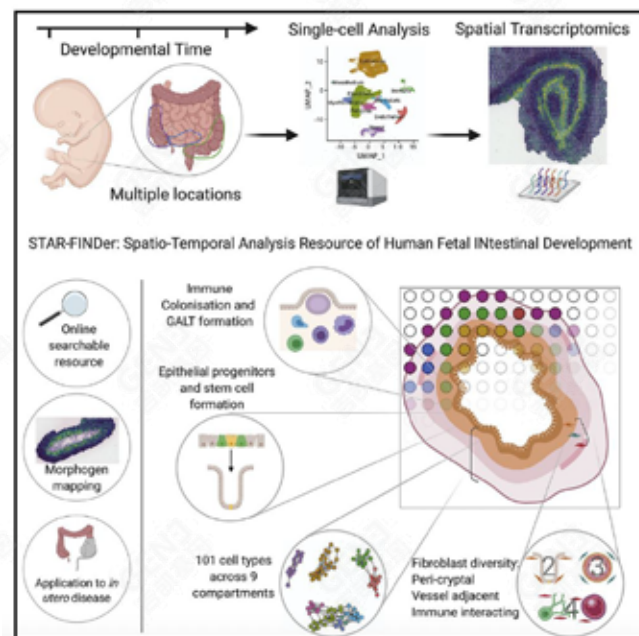
05.

单细胞水平人肠道发育时空分析

Cell

Spatiotemporal analysis of human intestinal development at single-cell resolution

Graphical Abstract



Authors

David Fawcner-Corbett, Agne Antanaviciute, Kaushal Parikh, ..., Paul R.V. Johnson, Hashem Koohy, Alison Simmons

Correspondence

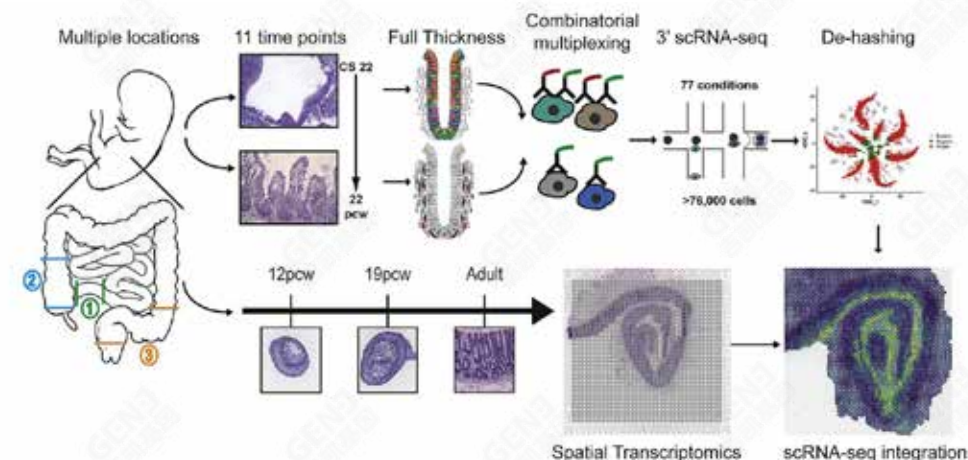
hashem.koohy@imm.ox.ac.uk (H.K.), alison.simmons@imm.ox.ac.uk (A.S.)

In Brief

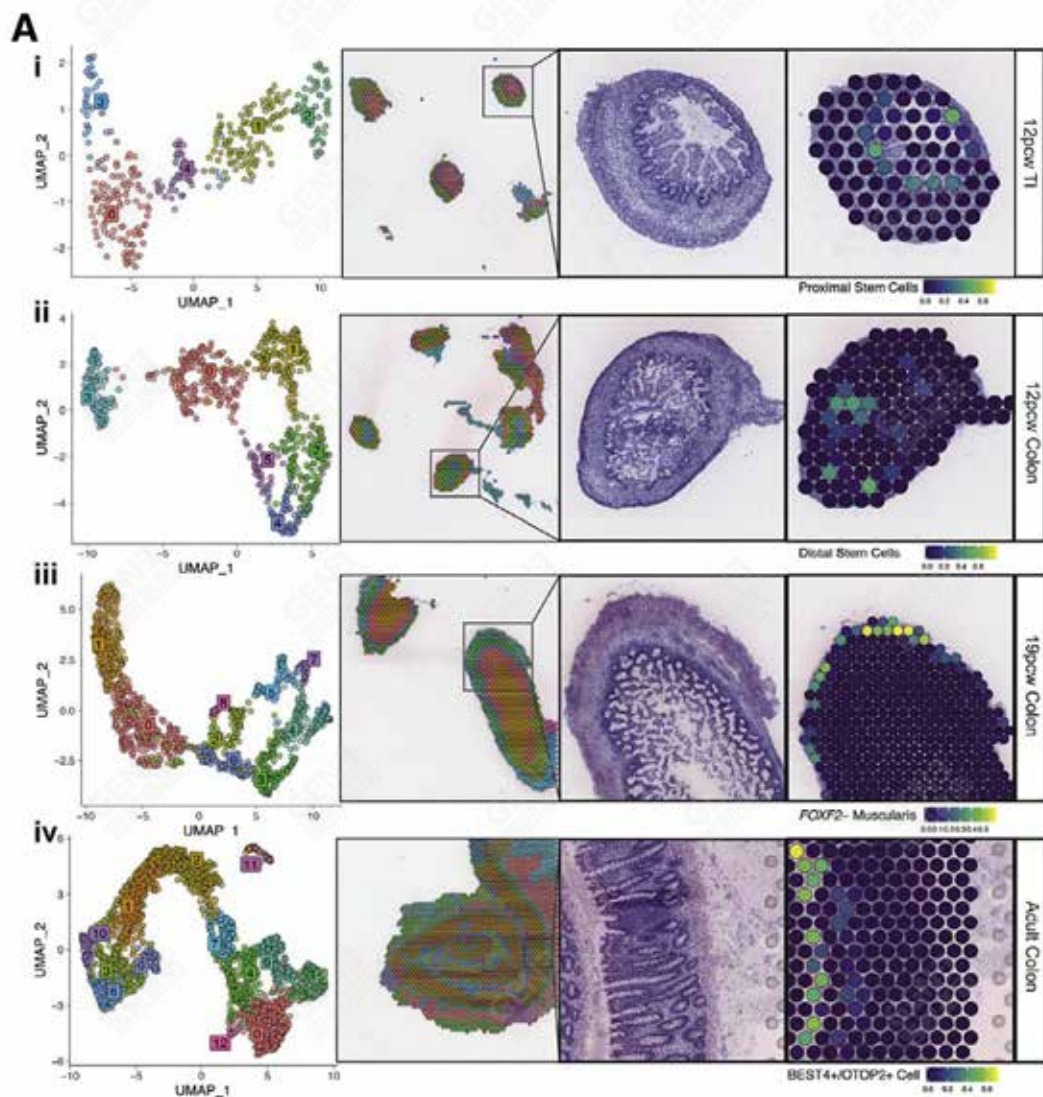
Fawcner-Corbett et al. chart human intestinal morphogenesis across time, location, and cellular compartments using a combination of single-cell RNA sequencing and spatial transcriptomics. The resulting online searchable atlas describes the principles of crypt-villus axis formation as well as neural, vascular, mesenchymal morphogenesis, and immune populations of the developing gut.

Resource

A

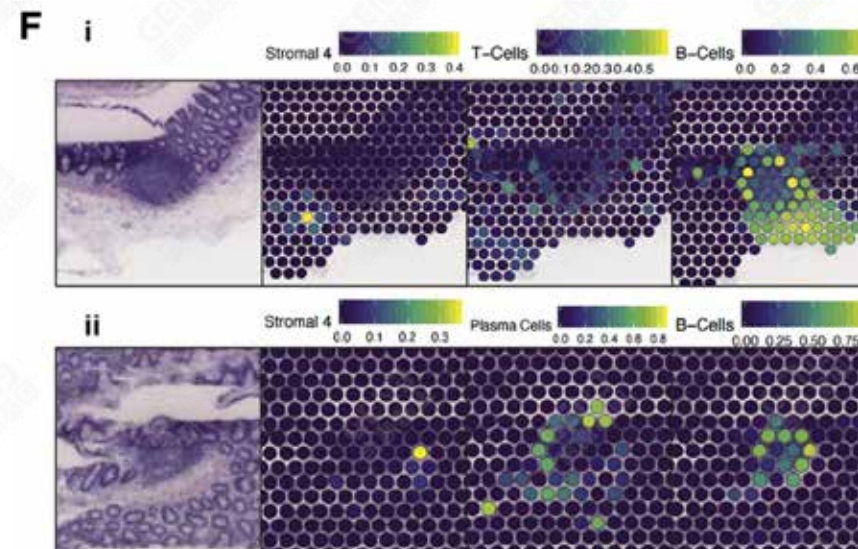
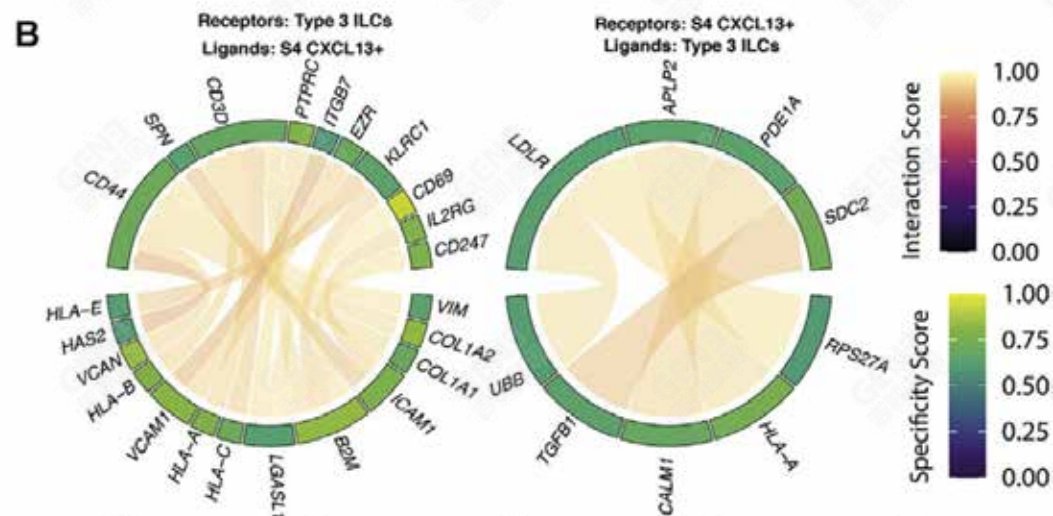


来自英国牛津大学的Fawcner-Corbett等人于2021年在*Cell*杂志上发表的一篇题为“*Spatiotemporal analysis of human intestinal development at single-cell resolution*”的文章，通过对多个部位和发育时间点的肠组织进行单细胞转录组与空间转录组测序，表征了肠道形态随时间的改。



- 四张切片分别对应孕后期12周（PCW）的回肠末端（TI），12pcw结肠、19pcw结肠，成人结肠。分别鉴定出5-13个空间spot cluster
- 使用单细胞数据的细胞类型信息映射定位了特征明显的细胞类型，如隐窝底部的肠道干细胞（ISC）（图A i和A ii），外肌的FOXF2-肌层细胞（图A iii），以及隐窝顶部的BEST4/OTOP2细胞（图A iv）

鉴定一类免疫滤泡相邻的成纤维细胞



- 细胞受体配体互作分析发现，S4成纤维细胞与不同免疫细胞的存在细胞通讯，例如通过VCAM1-ITGB7互作对与ILC3进行信号传递。这一结果在空间位置上得到证实
- 因此表明，S4细胞是成年结肠中免疫滤泡相邻的成纤维细胞

THANK YOU !!



扫描二维码，获取更多资料

2022-5-11

上海吉凯基因医学科技股份有限公司
SHANGHAI GENECHEM CO.,LTD.

GEN3
吉凯基因