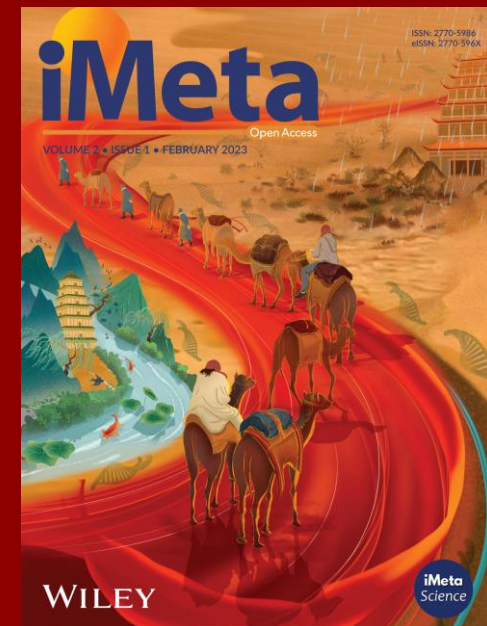




跨多种技术、器官和生物学重复的空间转录组学聚类方法基准测试

陈仁杰， 么玥， 钱竞扬， 彭昕*， 邵鑫*， 范晓辉*



浙江大学药学院、浙江三角智慧绿洲创新中心、现代中药创制全国重点实验室、浙江中医药大学附属宁波中医院

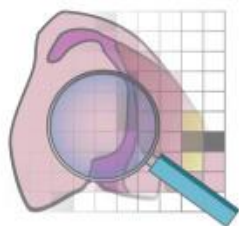
Renjie Chen, Yue Yao, Jingyang Qian, Xin Peng, Xin Shao, Xiaohui Fan. 2025. A comprehensive benchmarking for spatially resolved transcriptomics clustering methods across variable technologies, organs, and replicates. *iMeta* 4: e70084. <https://doi.org/10.1002/imt2.70084>.



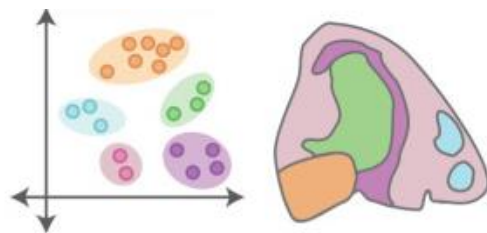
简介

空间聚类是空间转录组数据分析的关键步骤

空间转录组学数据



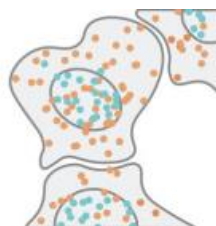
空间聚类



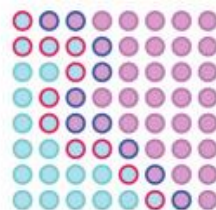
空间可变基因



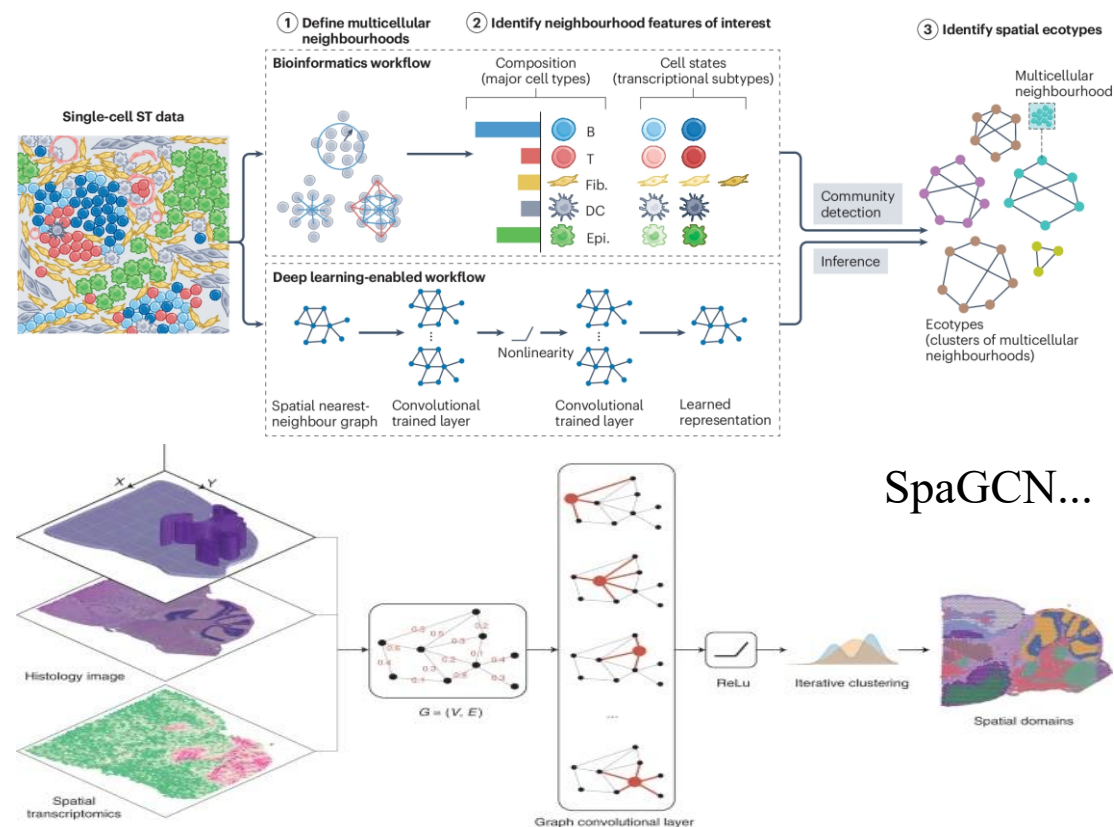
亚细胞分析



空间细胞通讯



已有大量空间聚类算法/工具

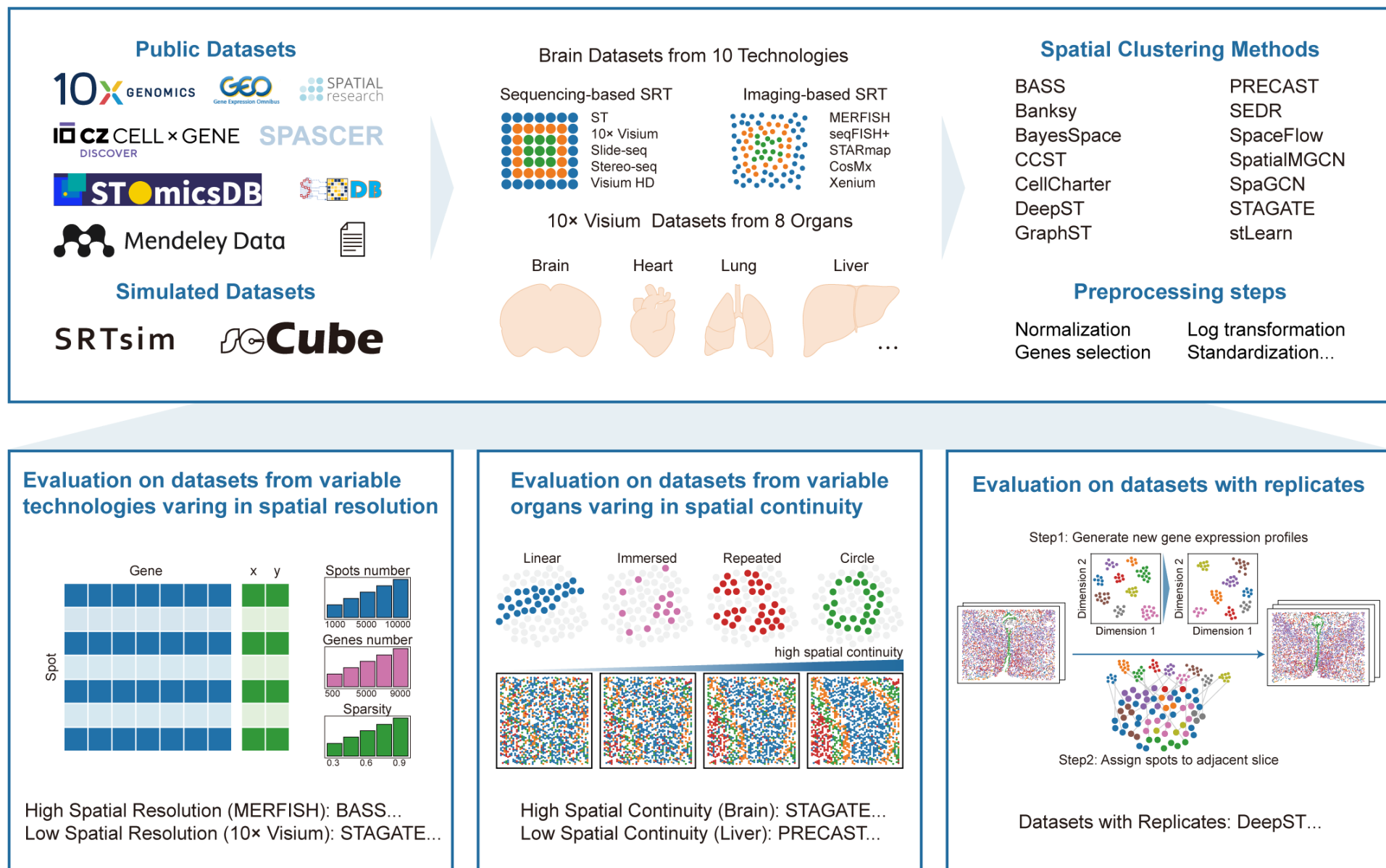


如何针对不同技术、器官、生物学重复的应用场景选择最合适的聚类算法？



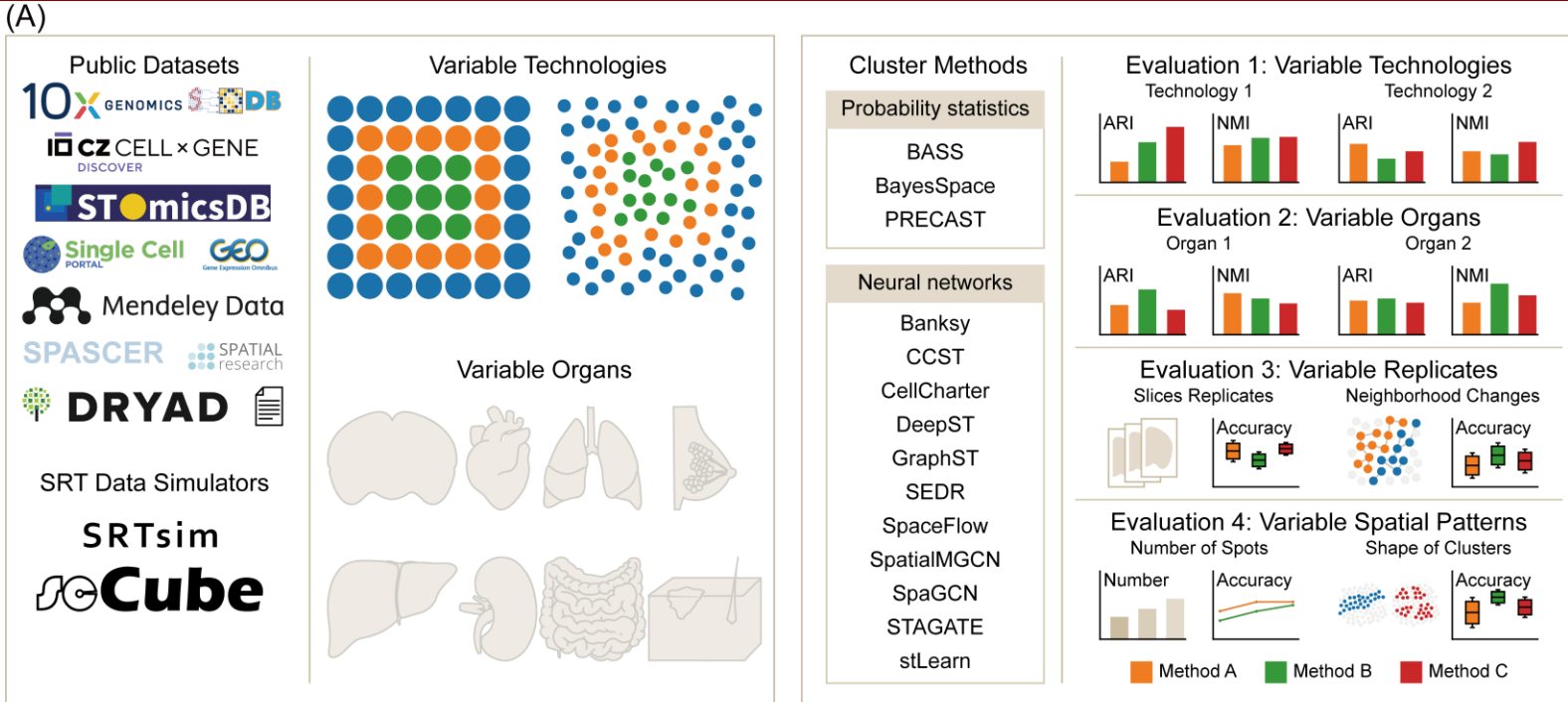
亮点

- 利用约600个数据集对14种空间聚类方法进行全面基准测试，涵盖十种技术和八种器官。
- 针对跨技术平台、器官类型和生物学重复等应用场景进行空间聚类方法推荐，涉及细胞类型聚类或空间域识别等聚类任务。
- 系统地评估了转录组数据特征和空间分布模式对方法聚类准确性的影响，并为空间聚类方法提供了优化的预处理流程。



空间转录组学数据聚类方法的基准测试框架

图1. 基准测试数据集和聚类方法概述示意图



(B)

Technology	Slices	Spatial Resolution	Spots per Slice	Genes Detected	Sparsity	Total Spots
Sequencing-based						
ST	35	200 μm I 100 μm	458	23371	0.822	16024
10 $\times$ Visium	151	100 μm I 55 μm	2446	32080	0.919	369357
Slide-seq	102	10 μm I	24569	23484	0.986	2506054
Stereo-seq	5	0.5 μm I 0.22 μm	72449	25077	0.970	362247
Visium HD	6	2 μm I	268419	18572	0.957	1610513
Imaging-based						
seqFISH+	14	$\leq 0.1 \mu\text{m}$ I	211	9961	0.819	2956
STARmap	5	0.2 ~ 0.3 μm I	1065	508	0.727	5327
MERFISH	41	$\leq 0.1 \mu\text{m}$ I	4882	182	0.631	200163
CosMx	6	$\leq 0.1 \mu\text{m}$ I	51700	957	0.779	310199
Xenium	3	$\leq 0.1 \mu\text{m}$ I	143211	291	0.751	429634

比较场景 1: 不同技术平台数据集的聚类方法比较

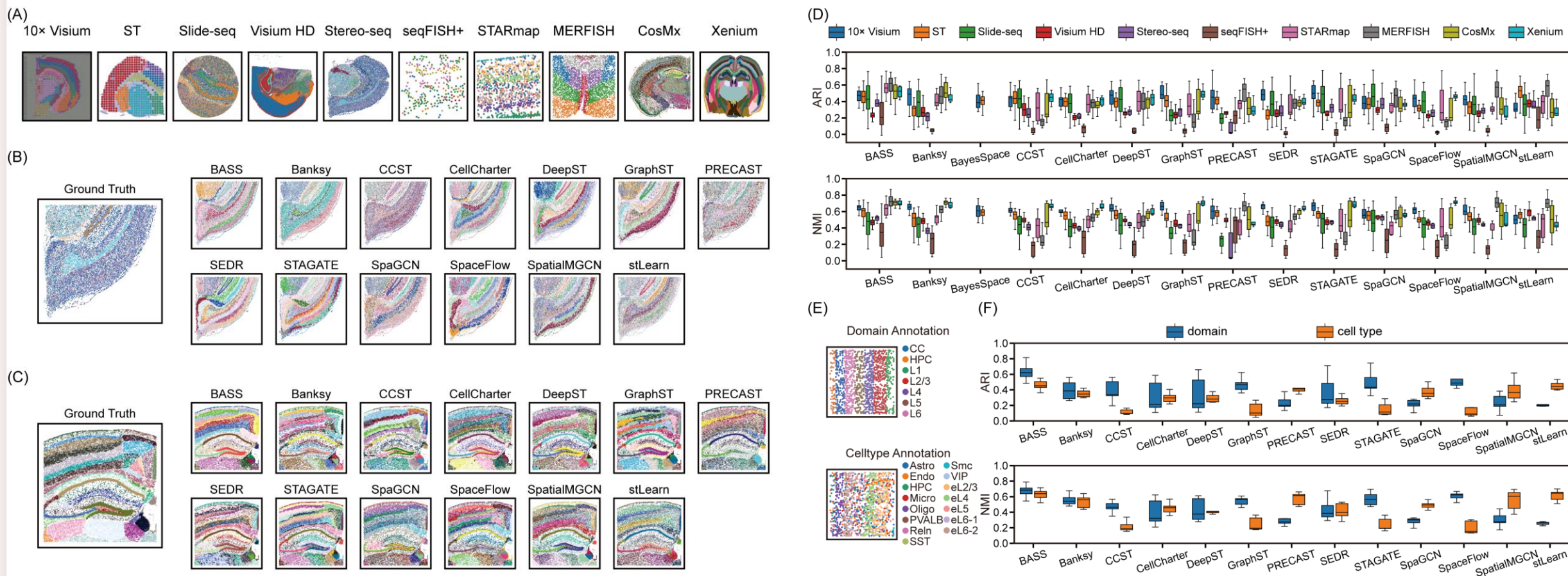


图2. 基于不同技术平台的空间转录组学数据的聚类性能比较

比较场景 2: 不同器官数据集的聚类方法比较

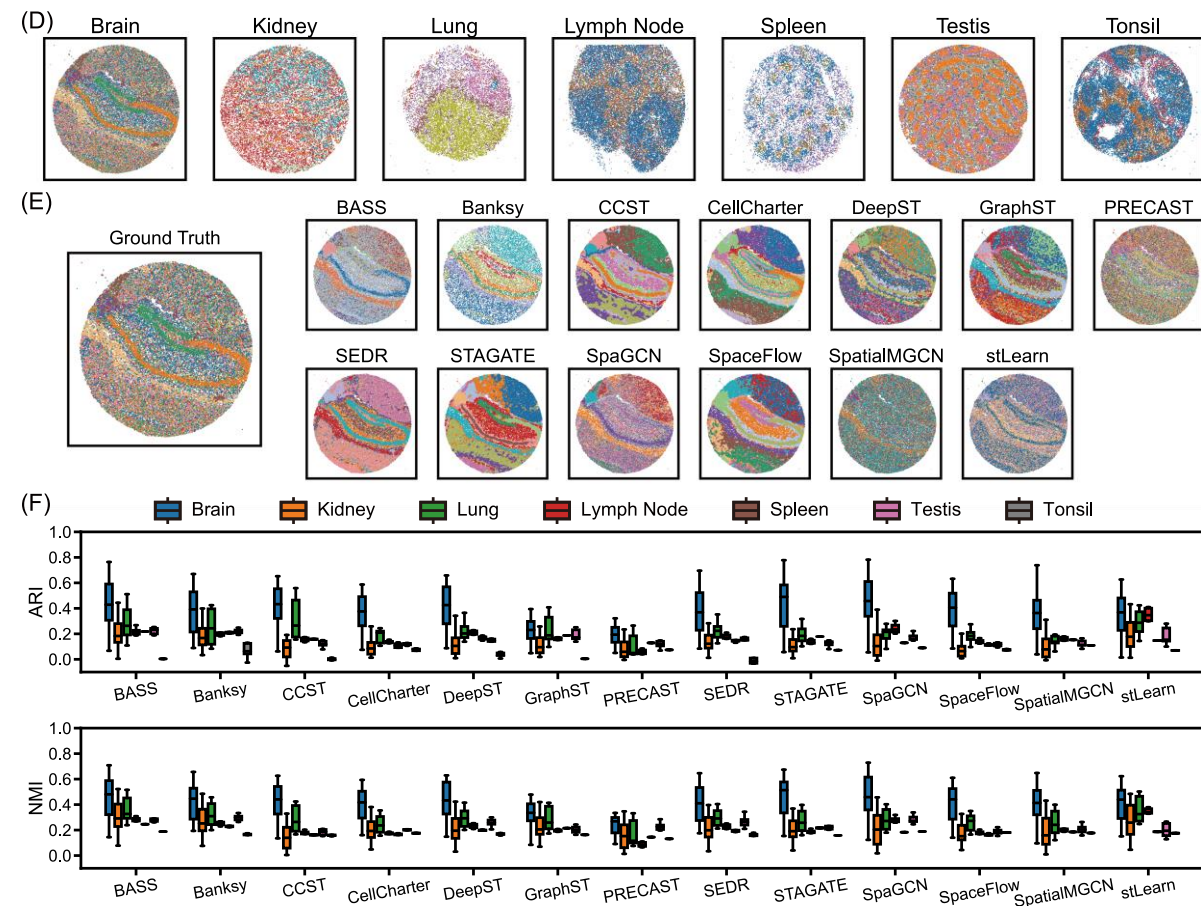
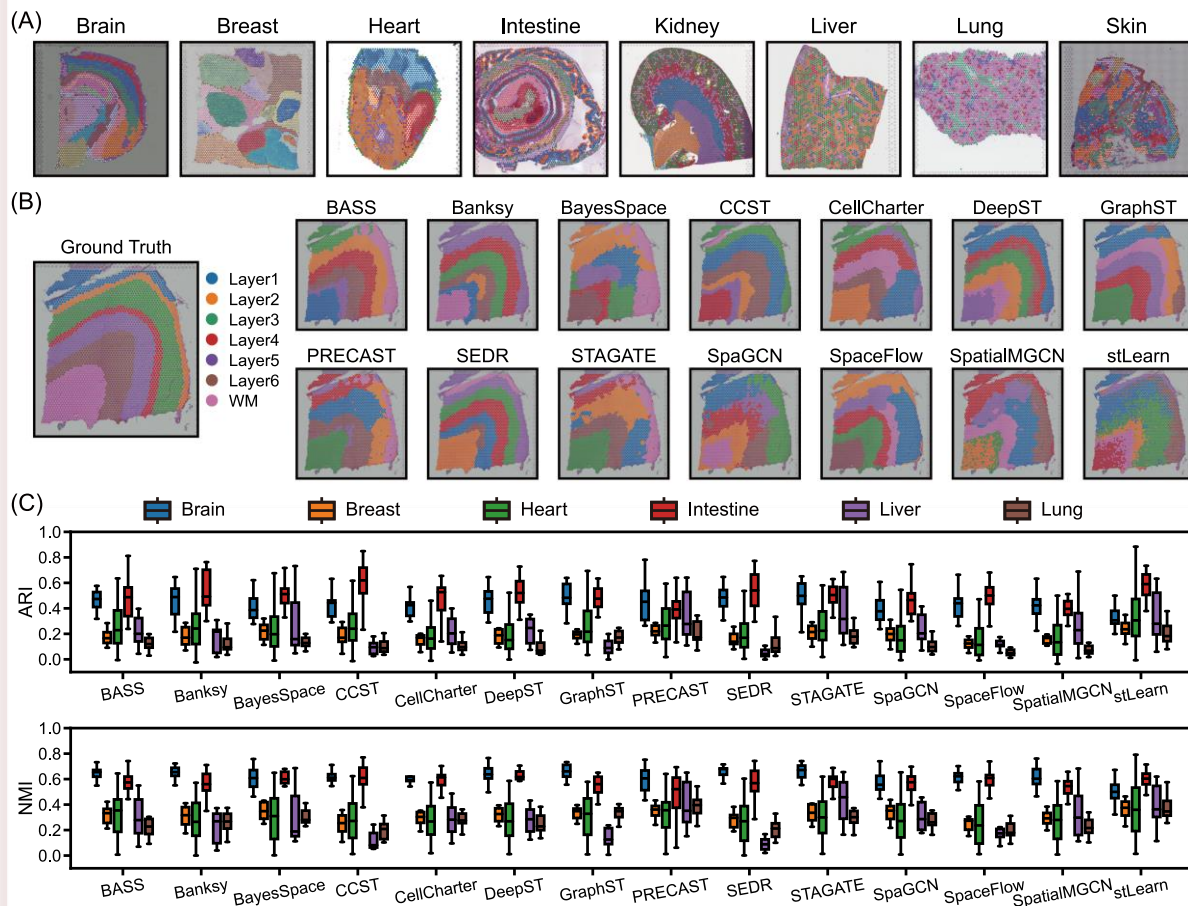


图3. 基于不同器官的空间转录组学数据的聚类性能比较

比较场景 3: 具有生物学重复数据集的聚类方法比较

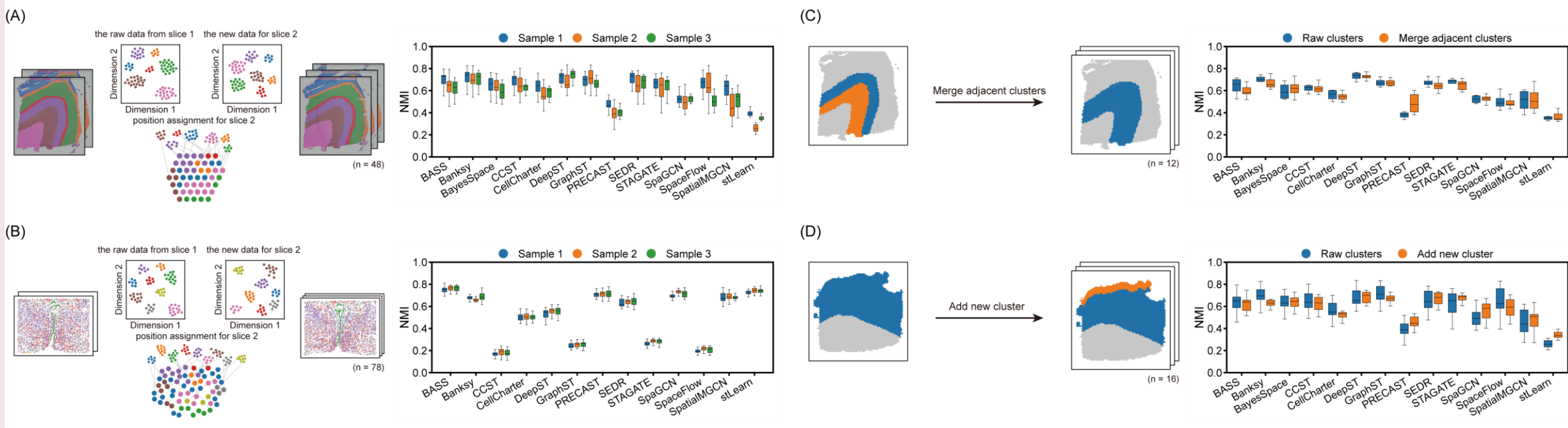


图4.基于具有生物学重复的空间转录组学数据的聚类性能比较

(A)利用人背外侧前额叶皮层10× Visium数据集模拟具有生物学重复的样本。(B) 基于小鼠下丘脑MERFISH数据集模拟具有生物学重复的样本。(C) 通过合并相邻空间区域模拟邻域改变的重复样本。(D) 通过增加新区域模拟邻域改变的重复样本。

箱线图展示了模拟数据集上的各个聚类方法NMI评分。

比较场景 4: 不同空间模式模拟数据集的聚类方法比较

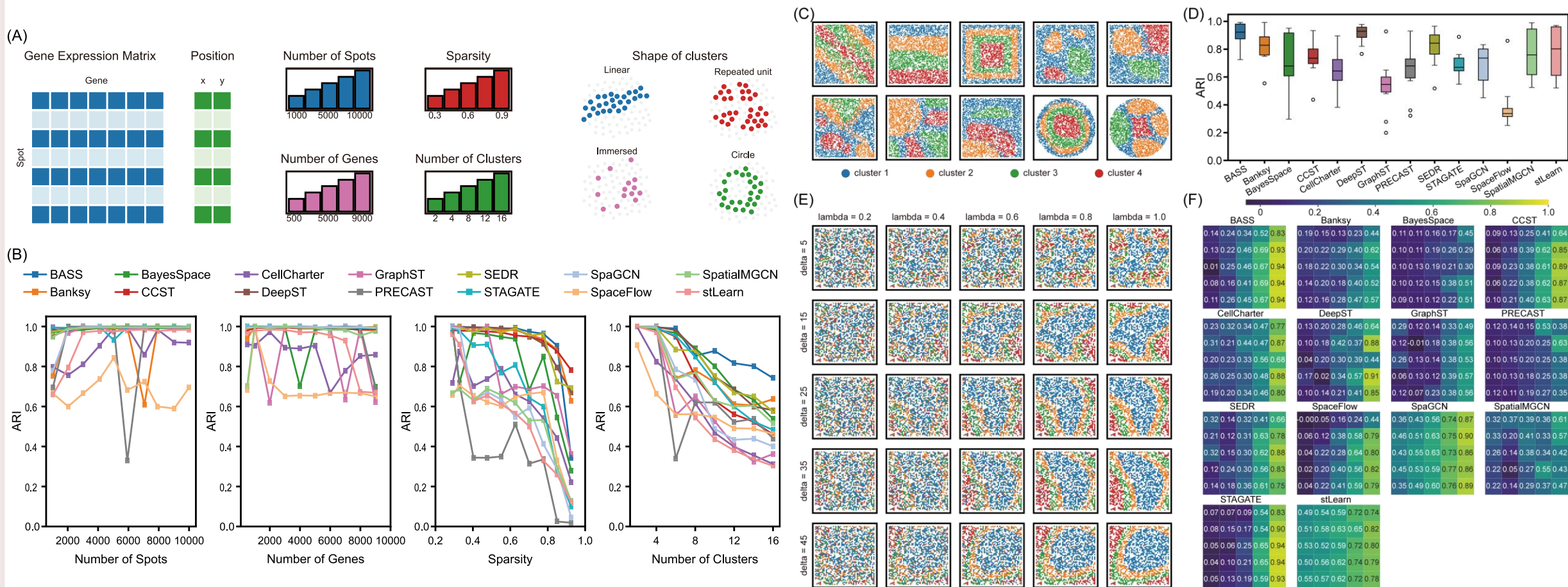


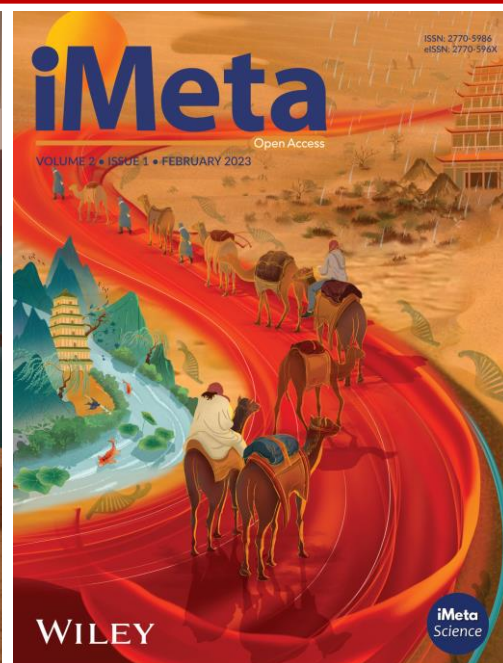
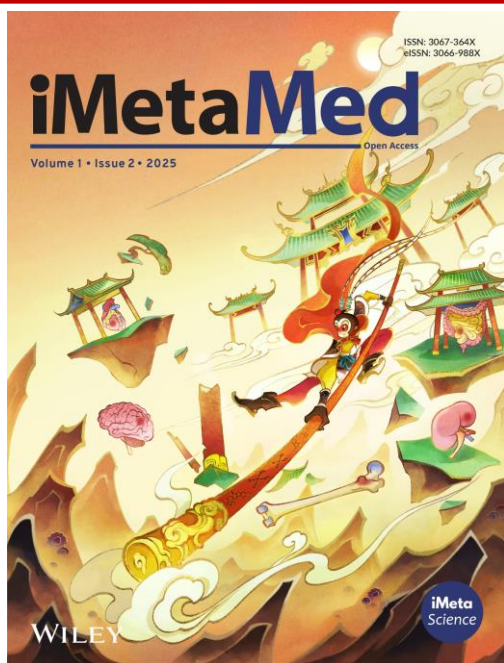
图5. 基于不同空间模式模拟数据集的聚类性能比较



总结

- ❑ 在这项研究中，我们对14种空间聚类方法进行了全面的基准测试，涵盖不同技术平台、器官类型和生物学重复，并针对具体的应用场景提供了方法推荐。
- ❑ 在不同应用场景中，方法性能差异主要受到算法模型设计和数据集自身特征影响，其中基因表达矩阵的稀疏性和空间域的分布特征是关键因素。
- ❑ 预处理策略同样会影响聚类性能，我们为每种方法提供了优化后的数据预处理流程。
- ❑ Github 网址： <https://github.com/ZJUFanLab/SRTBenchmark>

Renjie Chen, Yue Yao, Jingyang Qian, Xin Peng, Xin Shao, Xiaohui Fan. 2025. A comprehensive benchmarking for spatially resolved transcriptomics clustering methods across variable technologies, organs, and replicates. *iMeta* 4: e70084. <https://doi.org/10.1002/imt2.70084>.



iMeta(宏)期刊是由宏科学、千名华人科学家和威立共同出版，对标**Cell**的生物/医学类综合期刊，主编刘双江和傅静远教授，欢迎高影响力的研究、方法和综述投稿，重点关注生物技术、大数据和组学等前沿交叉学科。已被**SCIE**、**PubMed**等收录，最新IF 33.2，位列全球SCI期刊第65位(前千分之三)，中国第5位，微生物学研究类全球第一，中科院生物学双1区Top。外审平均21天，投稿至发表中位数87天。

子刊**iMetaOmics** (宏组学)、**iMetaMed** (宏医学)定位IF>10和15的生物、医学综合期刊，欢迎投稿！



主页: <http://www.imeta.science>

出版社: <https://wileyonlinelibrary.com/journal/imeta>

iMeta: <https://wiley.atyponrex.com/journal/IMT2>

投稿: iMetaOmics: <https://wiley.atyponrex.com/journal/IMO2>

iMetaMed: <https://wiley.atyponrex.com/journal/IMM3>



office@imeta.science

imetaomics@imeta.science



宣传片



iMeta



更新日期
2025/7/6