

DOI: 10.13376/j.cblls/2024184

文章编号: 1004-0374(2024)12-1572-16

空间转录组检测技术研究进展及展望

冯学超[#], 邵雨彤[#], 曹群发, 李佳豪, 张衙楠, 李文琰, 韩峻松, 张晓娜, 蒋敏*

(上海生物芯片有限公司, 生物芯片上海国家工程研究中心, 上海 201203)

摘要: 空间转录组技术是一种通过测序或成像技术, 在组织中进行基因空间表达检测的技术手段。该技术结合图像和建模分析算法, 可以获得基因表达和各种细胞类型在组织中的空间分布模式。本文总结了空间转录组的实验检测原理和数据分析方法。在实验方法上, 将目前的空间转录组技术分为三类, 并详细描述每种类别实验技术的原理和特征; 在数据分析上, 按照数据前处理、基础生物学信息挖掘、细胞注释和组织微环境特征分析的顺序, 描述其算法的原理和应用方向; 最后, 总结了空间转录组技术的局限性和未来发展方向, 希望能为基础医学和分子生物学的科研工作者提供帮助。

关键词: 空间组学; 原位检测; 基因表达; 数据挖掘

中图分类号: Q-33 文献标志码: A

Research progress and prospects of spatial transcriptomic technology

FENG Xue-Chao[#], SHAO Yu-Tong[#], CAO Qun-Fa, LI Jia-Hao, ZHANG Ya-Nan,

LI Wen-Yan, HAN Jun-Song, ZHANG Xiao-Na, JIANG Min*

(Shanghai Biochip Co., Ltd., National Engineering Center for Biochip at Shanghai, Shanghai 201203, China)

Abstract: Spatial transcriptomics provides spatially gene expression profiles in tissues via sequencing or imaging. Combined with algorithms for image analysis and modeling, spatial transcriptomics can offer the locations of genes and cell types in tissues. Here, we summarize the technical principles and data exploring methods of spatial transcriptomics. We divide the experiment technologies into three categories and describe their characteristics in detail and introduce the principle and application of various data exploring methods for pre-processing, basic bioinformation analysis, cell annotation and interpreting specific tissue microenvironment. Finally, we highlight the limitations and future development of spatial transcriptomics, hoping to provide assistance to researchers in basic medicine and molecular biology fields.

Key words: spatial transcriptome; *in situ* technology; gene expression; data mining

自从 2009 年 Tang 等^[1]首次实现单个小鼠卵裂球细胞全转录组 (mRNA) 的测序工作, 单细胞转录组测序技术 (single cell transcriptomic sequencing, scRNA-seq) 开始迅猛发展, 分子生物学研究迈入了全新的细胞尺度分辨时代。随着 scRNA-seq 的发展, 研究人员可以更深入了解单个细胞的基因表达模式, 揭示细胞类型和功能的多样性。然而, scRNA-seq 并不能提供空间信息。因此, 从组织原位探究细胞类型的空间分布及相互作用的研究就受到了限制。为了解决这一问题, 空间转录组检测技术应运而生。该技术可以在空间水平同时检测成千上万个

基因的表达信息, 并在保持细胞形态和位置信息的同时, 揭示不同细胞类型的分布和通讯关系。空间转录组检测技术为我们深入了解组织内细胞的空间分布和生物学过程提供了新的视角, 成为当前生命科学领域的热点技术之一。空间转录组检测技术也是众望所归, 被 *Nature Methods* 杂志评为 2020 年的年度技术^[2]。本文将对空间转录组检测技术的原

收稿日期: 2024-06-14; 修回日期: 2024-08-14

[#]共同第一作者

*通信作者: E-mail: jiangmin@shbiochip.com

理及数据分析方法进行阐述, 并讨论现阶段存在的问题和未来发展方向。

1 空间转录组检测技术研究进展

与传统转录组学技术相比, 空间转录组的难点在于如何在检测基因表达的同时保留基因的空间位置信息。为解决这一问题, 目前主要有三种策略: (1) 利用原位杂交技术直接检测目的基因的表达, 通过荧光成像或测序来确认不同基因的表达水平及分布; (2) 通过在测序文库中添加特殊 barcode 条码序列信息来标记基因表达的空间位置; (3) 利用显微切割或圈选感兴趣区域 (region of interest, ROI) 的方式, 通过 bulk RNA 测序方法分别获得组织上不同区域的整体基因表达信息。基于这三种策略的代表性技术流程图见图 1, 接下来, 本文将介绍采用这些策略的空间转录组检测技术。

1.1 基于原位杂交的空间转录组检测技术

1.1.1 原位杂交

原位杂交技术 (*in situ* hybridization, ISH) 由来已久。早在 20 世纪 60 年代, Gall 和 Pardue^[3] 就利用氘标记的核糖体 RNA 与非洲爪蟾 (*Xenopus laevis*) 卵母细胞中扩增的核糖体基因杂交, 通过放射自显影技术检测杂交体以确认核糖体 RNA 的扩增。然而, 该技术用到的放射性同位素危险性强。因此, 荧光原位杂交技术 (fluorescence *in situ* hybridization, FISH) 应运而生, 该技术利用荧光基因代替放射性同位素, 检测基因的原位表达^[4]。

1998 年, 单分子免疫荧光技术 (single molecular fluorescent *in situ* hybridization, smFISH) 横空出世。得益于探针技术的发展和数字成像显微镜的高分辨率特点, smFISH 可检测单个 RNA 分子, 从而更精准地定位和定量检测 RNA 的表达水平^[5]。然而, 由于荧光染料的光谱重叠, smFISH 的多路复用能力有限, 在创建转录组图谱时, 妨碍了有效的转录组水平的空间研究。为了编码尽可能多的转录本, 使用重复的 3~4 个彩色成像循环平台实现了顺序条形码方案, 并用超分辨率显微镜 (super-resolution microscopy, SRM) 记录颜色的空间位置顺序^[6]。但这个方法在没有 SRM 的情况下需要在一条转录本上杂交更多的探针以增强荧光信号, 这对转录本的长度有严格的要求。

2014 年, Lubeck 等^[7] 开发了连续 FISH 技术 (sequential fluorescence *in situ* hybridization, seqFISH), 将多轮杂交技术和颜色组合条形码技术相结合。每

轮杂交时, 任一转录本只会被一种颜色的探针杂交, 这样就避免了对显微镜和转录本长度的高要求。杂交轮数越多, 鉴定一个基因的颜色条形码就越长, 能够鉴定的基因数就越多。

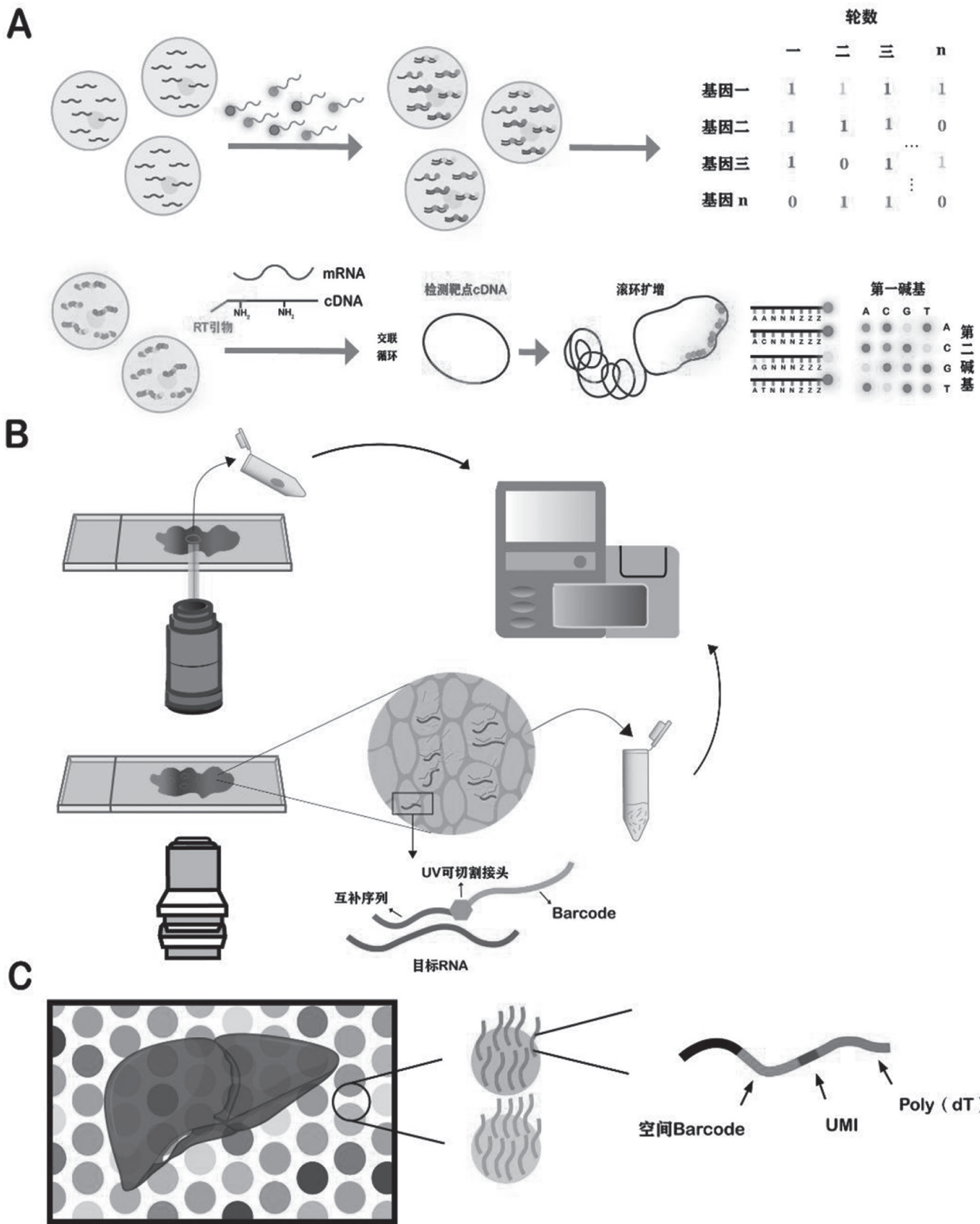
为了解决单分子荧光原位杂交的多重检测能力问题, Chen 等^[8] 开发了多重纠错 FISH 技术 (multiplexed error-robust fluorescence *in situ* hybridization, MERFISH), 用二进制条码 (位数代表杂交总轮数, 1 代表有荧光, 0 代表无荧光) 来定义靶标基因种类, 从而提高条形码的组合空间。

在增强检出信号的强度方面, 主流的策略有分支链 DNA 信号放大技术 (branched DNA, bDNA)、滚环扩增 (rolling circle amplification, RCA) 和杂交链式反应 (hybridization chain reaction, HCR)。这些方法都以非共价形式将大量荧光基因连接到探针上以放大信号, 提高信噪比。

bDNA 技术通过多级信号放大机制, 首先利用桥式探针与转录本特异性结合, 然后将主放大器序列结合在桥式探针上, 并将带有多个荧光基因的次放大器序列结合在主放大器序列上, 形成一个多分支簇, 这些分支可以结合大量的信号探针, 增强荧光信号强度^[9-11]。多级放大可能增加非特异性背景信号, 因此 RNAscope 对 bDNA 技术进行调整, 用两个 Z 型桥式探针杂交在转录本上, 只有两个探针都成功杂交上才能让主放大器成功连接。这种改良可以降低非特异性杂交的背景噪声, 适合 RNA 降解的样品, 如石蜡包埋组织 (FFPE)^[12]。

RCA 技术与挂锁探针技术的联合使用能够显著提高检测的灵敏度和特异性, 尤其适用于检测低丰度的核酸分子和在复杂生物样本中实现高分辨率的空间检测。挂锁探针的两端必须与目标转录本杂交且没有错配, 才能在连接酶的作用下成环。接下来, 通过滚环复制 (rolling circle replication, RCR) 可以将单一的圆形 DNA 模板持续扩展生成链 DNA 产物, 即可获得大量的探针拷贝数, 从而在原位放大信号^[13]。但滚环产物的长度可能导致实验结果的复杂性, 需要优化检测探针和标记方案以获得最佳的信号检测效果。

HCR 技术使用 2 种带荧光标记的发夹型探针进行循环杂交, 2 种探针都具有长茎和短茎, 当目标分子存在时, 探针会发生自组装形成带有大量荧光标记的长双链 DNA 聚合物, 放大信号^[14]。此过程不需要酶参与, 因此反应条件较为温和。相比于 RCA, HCR 的放大倍数通常较低, 不适合于需要



(A)基于原位杂交、原位测序的空间转录组检测技术。(B)基于空间barcode条码的空间转录组测序技术。(C)基于特定区域的空间转录组检测技术。

图1 空间转录组检测代表性技术原理图

极高灵敏度的检测。

以上是 ISH 技术的诞生及其衍生方法的发展过

程。这些技术利用荧光信号靶向不同的转录本，通过结合现代探针设计和荧光标记技术，大大提高了

基因和 RNA 的检测灵敏度和特异性。特别是信号增强策略的应用, 进一步提高了检测的信号强度和精确度。尽管各有优缺点, 这些技术为研究基因表达的空间分布及其动态变化提供了宝贵的工具。

1.1.2 原位测序

除了利用荧光来表征转录本外, 原位测序技术通过将 RNA 原位转录为 cDNA, 再进行测序, 提供 RNA 的空间分布信息。该技术结合了成像和测序技术的优点, 以在组织中实现 RNA 的高通量、高精度检测。该方法先在原位将目标转录本逆转录为环化 cDNA 或与挂锁探针杂交, 再通过滚环扩增大量复制以进行后续测序。

测序时, 可通过直接检测 cDNA 序列以鉴定基因, 如荧光原位测序 (fluorescent *in situ* sequencing, FISSEQ) 等^[15-16], 或者检测挂锁探针上为不同基因设计的特异性 barcode 序列, 间接鉴定基因如 ISS (*in situ* sequencing) 等^[17]。这种方法摆脱了荧光的颜色数量限制, 与非多轮 ISH 技术相比, 提高了检测的基因数量。但是, 受限于测序技术读长较短、检测时间较长、核酸扩增产物紧密堆积、转录物检测效率较低, 仅能检测到约 0.2%~1% 的转录物^[18]。

为了改善这些问题, 研究者们进行了优化。Alon 等^[19]将膨胀显微成像技术与 FISSEQ 技术相结合, 开发了 ExSeq (expansion sequencing) 技术, 该技术通过将转录本合并到聚电解质凝胶中并使其扩展, 避免了核酸扩增产物的堆积, 提高了高丰度基因的检测率, 其检测效率和 rRNA 的 reads 比例与 bulk RNA 测序相近, 提供了准确的 RNA 表达水平数据。

Wang 等^[20]开发出了 STARmap (spatially-resolved transcript amplicon readout mapping) 技术, 利用分子内连接特异性扩增核酸 (specific amplification of nucleic acids via intramolecular ligation, SNAIL) 的挂锁探针设计, 提高了检测的特异性和准确性, 并通过扩增物交联成水凝胶, 减少了背景噪音和自发荧光的干扰。该技术绕过了逆转录步骤, 解决了逆转录和测序错误导致检测效率和准确性低的问题。通过将扩增物交联成水凝胶并清除组织中的蛋白质和脂类, 减少了背景自发荧光的干扰。但该技术需要精确设计特异性 barcode 和探针序列, 这增加了实验的复杂性和成本。

随着原位测序技术的不断发展和改进, 新技术如 ExSeq 和 STARmap 已经展示了应对检测效率低、测序错误等挑战的突破性进展, 进一步提高了基因

检测的精度和信噪比。在未来, 随着测序技术的进一步优化, 有望在读长和检测时间上得到改善, 从而提高转录物检测效率, 为疾病研究和精准医疗提供更强大的工具。

1.2 基于空间barcode条码的空间转录组测序技术

在空间转录组研究中, 一些技术在高通量测序文库序列中增添空间 barcode 条码, 通过读取该序列信息来确认基因转录本在组织切片上的空间表达位置。与原位杂交策略在空间原位进行荧光检测或测序相比, 空间 barcode 条码策略可检测的基因通量更高。但是, 与原位杂交及成像策略相比, 其空间检测分辨率更低。目前, 使用 barcode 条码的空间转录组检测技术主要有 ST (spatial transcriptomics)、Visium (10X Visium)、Slide-seq、HDST (high-definition spatial transcriptomics)、DBiT-seq (deterministic barcoding in tissue for spatial omics sequencing)、Stereo-seq (spatial enhanced resolution omics-sequencing)、Pixel-seq、Slide-tag 等。

在 Ståhl 等^[21]开发的 ST 技术中, 组织切片平铺在芯片上并被透化, 使得组织切片中的转录本进入 spots 内并被检测探针捕获, 用于建库和测序分析。每张组织切片的最大检测面积为 $6.2 \times 6.6 \text{ mm}^2$, 在这个面积内均匀分布了 1 007 个圆形 spots 捕获区域, 每个 spot 内锚定了大量的检测探针, 每条探针都包含记录空间位置的 barcode 条码、定量 UMI 条码和捕获转录本 3' 末端的 ployT 序列。

在 10X Genomics 公司优化的 Visium 技术中, 每张切片的检测面积扩大至 $6.5 \times 6.5 \text{ mm}^2$, 甚至 $11 \times 11 \text{ mm}^2$, spot 直径缩小为 $55 \mu\text{m}$, spots 之间的圆心间距缩小至 $100 \mu\text{m}$ 。检测区域的扩大, 搭配 spot 直径和间隔的缩小, 不仅增加了 spots 的总覆盖面积, 也显著提高了检测分辨率。然而, $55 \mu\text{m}$ 超过了一般细胞的直径大小, 导致一个 spot 中可能捕获到几个到几十个细胞的转录本信息, 分辨率低。

为此, Rodrigues 等^[22]开发出 Slide-seq 技术, 该技术使用了与 Drop-seq 技术^[23]中类似的 beads, beads 上带有独特的空间 barcode 条码标签和捕获转录本的探针。这些 beads 被随机分配到玻片表面, 然后对 barcode 条码原位测序以确认每个 bead 的空间位置。接下来与 Visium 技术类似, 将组织切片平铺在 beads 阵列的表面并进行透化, 使得组织切片中的转录本被 beads 上的探针捕获, 用于建库和测序。由于 beads 的直径为 $10 \mu\text{m}$, 比 Visium 具有更高的分辨率。

Slide-tags 建立在 Slide-seq 方法的基础上, 改进了 Slide-seq 无法获得单细胞分辨率的问题^[24]。该技术使用带有多个相同 DNA 条形码的特殊珠子 (beads) 阵列, 以标记每个 bead 在阵列中的位置。当组织切片被转移到这些 beads 上时, 条形码渗入细胞核。通过检测每个细胞核中的不同条形码含量, 可以精确定位细胞核的位置。此方法实现了单个细胞核的定位和 RNA 表达, 即使在固定后的组织中也能有效应用, 从而重建组织中的细胞模式。

Vickovic 等^[25]开发了 HDST (high-definition spatial transcriptomics) 技术, 其原理与 Slide-seq 类似, 但是用多轮 FISH 代替原位测序以确定 barcode 序列及 beads 的空间位置, beads 的直径缩小到 2 μm , 达到了亚细胞水平的分辨率。但 Slide-seq 和 HDST 技术定位 barcode 的效率低且操作过程复杂。

针对这一点, Liu 等^[26]开发出 DBiT-seq (deterministic barcoding in tissue for spatial omics sequencing) 技术, 分辨率可达 10 μm 。该技术与上述 barcode 条码技术最大的不同之处是取消了从组织切片中释放转录本的透化操作, 取而代之的是将 barcode 条码探针通过微流控技术导入到组织切片中与转录本结合。此外, 与 ST 和 Visium 技术相比, DBiT-seq 的单位面积检测到的基因和 UMI 数量提高 2 倍左右。

Chen 等^[27]开发出 Stereo-seq (spatial enhanced resolution omics-sequencing) 技术, 将 spot 的直径缩小到 220 nm, 进一步提高了使用 barcode 条码的空间转录组技术的检测分辨率。

Pixel-seq 通过高密度 DNA 条形码簇实现转录本的原位捕获^[28]。先前的同类技术如 Stereo-seq, 以及更早期的 Slide-seq 和 HDST, 大多需要随机散布 DNA 点阵。而 DNA 芯片活字印刷技术 (polony gel) 将带有不同空间条形码信息的 DNA 簇和相应的引物共价交联在聚丙烯酰胺薄膜表面, 这些 DNA 簇的大小可以控制在约 1 μm 。通过限制转录模板在芯片表面的扩散, polony gel 芯片有效提高了空间分辨率。当前, 通过引入 barcode 来实现基因转录本在组织切片中的空间定位的检测技术均通过不同的方法提高检测的分辨率和效率, 但捕获效率及分辨率仍不及基于序列原位杂交和原位测序的空间组学方法。未来的发展方向包括进一步提高分辨率和检测效率, 尝试将 barcode 与基于原位杂交和原位测序等的高分辨率空间组学方法联合使用, 还原更高分辨率的组织单细胞空间分子图谱。

1.3 基于特定区域捕获的空间转录组检测技术

与原位杂交和 barcode 条码技术相比, 利用物理切割等技术将 ROI 从整体分离后再进行转录组检测的方法要简便得多, 能直接靶向关键亚结构区域, 获得的转录组数据也更贴合研究目的。但是, 针对 ROI 进行 bulk RNA 测序会丢失基因在 ROI 中确切的表达位置, 使得该策略的检测分辨率取决于 ROI 的大小。这种策略的经典技术方案是将激光捕获显微切割技术 (laser capture microdissection, LCM) 和 RNA 建库测序联合使用。LCM 可在保证捕获区域和周围组织区域完整的情况下, 从组织切片中获得 ROI 组织, 是迄今为止最常用的显微解剖方法^[29]。LCM 的激光主要使用红外光 (infrared ray, IR) 和紫外光 (ultraviolet radiation, UV)。在早期, LCM 主要使用 IR 将 ROI 区域短暂加热至 90 $^{\circ}\text{C}$, 使得一种特殊材质的薄膜熔化并与 ROI 融合, 从而实现捕获 ROI 的目的。然而, 高温不仅会对组织造成一定程度的破坏, 还会影响 RNA 的完整性^[30]。因此, UV 光源逐渐进入研究者的视野。与 IR 相比, UV 照射 ROI 的边缘即可与周边组织分割开, ROI 区域加热程度较低, 配合激光压力弹射器或重力将 ROI 区域组织转移到收集器中, 避免物理接触, 从而防止交叉污染。由于切片组织中的 RNA 含量低, 因此需要适合微量 RNA 的扩增建库方法, 比如 CEL-seq (cell expression by linear amplification and sequencing)、Smart-seq (switching mechanism at 5' end of the RNA transcript)^[31-33] 等。

另一种基于物理切割的技术是 Tomo (tomography-array/seq), 可以获得三维空间转录组表达图谱。该方法分别对三个相同组织块从三个方向进行连续切片, 然后利用表达谱芯片或测序技术单独检测每张切片的转录组并整合成一个整体。这种基于切片数据重建组织 3D 表达谱策略的方法已在小鼠脑、果蝇胚胎、斑马鱼胚胎等组织中得到应用^[34-36]。

除了物理切割外, 也有一些无须切割组织, 而是利用光学或分子标记从 ROI 区域直接捕获转录本或细胞的方法。转录组体内分析 (transcriptome *in vivo* analysis, TIVA) 是基于这种策略的第一种方法, 其原理是通过细胞穿透肽将 polyU 序列探针导入活细胞中, 此时的 polyU 与两段 polyA 序列互补结合, 而两段 polyA 之间及其与 polyU 序列间都与光激活生物素结合。被 405 nm 的激光激活后, ROI 内细胞中的 polyU 探针暴露出来并与带有 polyA 的转录本结合, 完成捕获^[37]。该技术的最大优势是可对活

体组织进行实验, 但其通量极低, 一次只能分析十几个细胞。

另一种应用光学的技术是 NanoString 公司推出的 GeoMx digital spatial profiling (DSP) 技术。该方法可在一张切片上完成荧光染色成像、ROI 选择和高通量转录组分析^[38]。DSP 技术使用特殊设计的探针, 其包含与目标基因转录本特异性结合的序列及为该基因设计的独一无二的 barcode 核酸条码, 二者通过 UV 可切割接头 (photocleavable linker, PC) 连接。实验时, 将荧光抗体与探针一起杂交到组织切片中, 然后通过荧光成像技术选择 ROI 区域, 用 UV 照射 ROI 区域, 使得 UV 可切割接头从而释放 barcode 条码, 然后收集这些游离的 barcode 核酸用于建库测序, 通过分析 barcode 条码即可获得基因的表达信息。DSP 技术能检测的基因数量取决于选择的检测 panel。目前, DSP 技术成熟的转录组检测 panel 有肿瘤基因图谱和全转录组图谱, 前者包含 1 800 个左右的肿瘤基因, 后者几乎囊括了人、小鼠所有的蛋白质编码基因, 已在神经、肿瘤和传染性疾病等研究中得到广泛应用^[39-41]。

这些技术在检测精度和效率上不断进步, 但仍存在如检测通量低、实验复杂度高和成本昂贵等挑战。未来的研究应致力于提高检测效率、降低成本, 同时进一步优化数据整合和分析方法, 为基础研究、药物开发和临床应用提供更强有力的工具。

1.4 空间转录组技术新时代

使用上述三种策略的空间转录组检测技术普遍存在以下问题中的一种或几种: (1) 实验操作繁琐; (2) 能够检测到的基因靶标数量少; (3) 检测分辨率低, 无法达到单个细胞的分辨率, 或者无法进行全目标区域检测。然而, 随着空间转录组研究的不断深入, 科研工作者们需要能够在组织中进行“空间高通量单细胞转录组检测”的技术。10X Genomics 和 Nanostring 为此开发了新的单细胞分辨率空间转录组检测技术: Xenium 和 CosMx spatial molecular imager (SMI)。

Xenium 是一种原位测序技术, 其利用滚环扩增来增强组织切片内靶 RNA 的信号, 并能保持组织完整性。为了扩大检测 RNA 靶标数量, 该技术选择了多轮杂交策略。此外, 图像 AI 分析软件会根据细胞核染色图像对扫描区域进行细胞分割, 从而获得空间“单细胞”转录组信息^[42]。

SMI 技术是一种多轮 ISH 技术, 同样使用图像 AI 分析软件进行细胞分割。其探针设计基于 MERFISH

设计方法, 可实现较低 RNA 检出错误率。此外, 该技术使用了 DSP 技术中的 PC 接头结构, 提高了检测结果的信噪比^[43]。虽然这两种原位检测技术都能获得空间“单细胞”转录组信息, 但是 Xenium 技术允许在检测 panel 中添加超过 100 个自定义 RNA 靶标, 而 SMI 技术的检测靶标总数多于 Xenium, 可达到 1 000 个 RNA 靶标。在 2024 年上半, 靶标数量增加到 6 000 个, 相信未来能够实现全转录组 (whole transcriptome atlas, WTA) 检测 (预计 2025 年实现)。此外, SMI 因为在荧光图像中用膜蛋白 marker 标记了细胞膜结构, 因此比 Xenium 技术具有更高的细胞识别准确性。

上述基于三种策略的部分空间转录组检测技术方法的汇总见表 1。未来, 随着空间转录组技术的不断发展和创新, 我们可以期待更高精度和更大通量的检测方法出现。

1.5 空间多组学技术

系统生物学是一个复杂的过程, 通过单一组学来解释生物学问题是片面的。空间多组学技术通过将单细胞组学、代谢组学、蛋白质组学、表观遗传学等与空间转录组学相结合, 可实现一次性获取细胞的空间位置以及同一细胞内的多个组学信息。先进的空间多组学技术不仅能够帮助识别新的细胞亚型, 还能够在特定的空间分辨率下观察细胞内和细胞间的分子相互作用。这些技术能够从染色质结构到细胞表型, 从 DNA 到代谢物的多个层次获取数据, 通过多组学方法综合基因组、表观组、蛋白质组和代谢组的信息, 为我们提供了更全面的视角以理解生物系统中的机制关系。

通过将空间转录组与 scRNA-seq 数据联系起来, 可以识别新的细胞类型, 并且可以从细胞类型共定位模式中推断细胞间相互作用, 用于研究癌症组织的微环境以及剖析肿瘤异质性等, 其持续发展将重塑癌症研究前景, 并为癌症的精确预后和治疗开辟新的可能性。此外, 由中国生物医药与健康研究院研发的 MISAR-seq (microfluidic indexing-based spatial assay for ATAC and RNA-sequencing) 空间多组学技术^[44], 通过微流控芯片依赖的靶向 barcode 递送系统, 在保留细胞空间位置信息的前提下同时实现了细胞内 ATAC 和 RNA 两种组学信息的捕获, 能够在构建时空多组学图谱的同时揭示启动子及增强子形成的顺式调控元件的调控机制。将空间转录组学和 TCR 测序相结合^[45], 对于追踪和解剖转移癌组织中的 T 细胞浸润及其相互作用至关重要。

表1 空间转录组检测技术列表

方法名称	发表年份	空间信息获得方式	分辨率	等级	参考文献
smFISH	1998	原位杂交	单分子	亚细胞	[5]
SRM	2012	原位杂交	单分子	亚细胞	[6]
RNAscope	2012	原位杂交	单分子	亚细胞	[12]
bDNA-smFISH	2013	原位杂交	单分子	亚细胞	[11]
seqFISH	2014	原位杂交	单分子	亚细胞	[7]
MERFISH	2015	原位杂交	单分子	亚细胞	[8]
CosMx SMI	2022	原位杂交	单分子	亚细胞	[43]
ISS	2013	原位测序	单分子	亚细胞	[17]
FISSEQ	2014	原位测序	单分子	亚细胞	[15]
STARmap	2018	原位测序	单分子	亚细胞	[20]
ExSeq	2021	原位测序	单分子	亚细胞	[19]
Xenium	2023	原位测序	单分子	亚细胞	[42]
ST	2016	空间barcode条码	100微米	多细胞	[21]
Slide-seq	2019	空间barcode条码	10微米	单细胞	[23]
Slide-tags	2023	空间barcode条码	单核	单细胞	[24]
HDST	2019	空间barcode条码	2微米	亚细胞	[25]
Visium	2020	空间barcode条码	55微米	多细胞	[21]
DBiT-seq	2020	空间barcode条码	10微米	单细胞	[26]
Stereo-seq	2022	空间barcode条码	220纳米	亚细胞	[27]
Pixel-seq	2022	空间barcode条码	1微米	亚细胞	[28]
LCM-seq	2016	选择特定区域	ROI	根据ROI大小可变	[29]
Tomo-array	2012	选择特定区域	根据切片厚度可变	组织器官3D结构	[35]
Tomo-seq	2013	选择特定区域	根据切片厚度可变	组织器官3D结构	[36]
TIVA	2014	选择特定区域	ROI	根据ROI大小可变	[37]
GeoMx DSP	2020	选择特定区域	ROI	根据ROI大小可变	[38]

空间多组学被 *Nature Methods* 杂志评选为 2023 年度最值得期待的技术之一^[46]。通过联合多维度的数据，空间多组学能够揭示细胞功能所需的关键分子作用，揭示疾病的分子机制和细胞特异性变化。空间多组学为理解复杂疾病的病理机制和开发新型治疗策略提供宝贵的参考资料，为精准医疗方案的开发提供支持，从而促进个性化治疗的实施和健康管理的可持续发展。

2 空间转录组数据分析策略

空间转录组数据分析的目标是整合基因表达及其空间位置的信息，以便从数据中挖掘有价值的生物信息。数据分析的流程可分为原始数据前处理、基础生物学信息挖掘、细胞注释、组织微环境特征分析(图2)。

目前有一些共享的空间转录组数据库可以提供给研究人员使用，如 STOmicsDB (Spatial Tran Omics DataBase) 数据库、SPASCER 数据库、SpatialDB (Spatial DataBase) 数据库等，涵盖了多达 17 个物种

的上百个数据集。

2.1 原始数据前处理

不同空间转录组技术获得的原始检测数据，需要根据数据类型选择合适的分析工具进行前处理，以便于进一步的下游分析。对于包含荧光成像图片数据的 smFISH 和原位测序技术，需要检出转录本在切片中的表达位点、转录本的种类，以及对图像进行分割以获得细胞分布结果，从而获得细胞-基因表达矩阵用于后续分析。在这个前处理过程中可供选择的方法有 Starfish^[47]、SMART-Q (Single-Molecule Automatic RNA Transcription Quantification)^[48]、dotdotdot^[49] 等。通过如阈值处理、基于翻译的图像配准、高帽滤波、高斯拉普拉斯算子、分水岭算法等经典图像处理方法实现图像配准、spots 及 barcode 检出、细胞分割等目的。Xenium 和 CosMx SMI 有专门的数据前处理软件 Xenium Ranger 和 AtoMxTM SIP (Spatial Informatics Platform)。对于使用 NGS 技术获得的数据，可以参考传统的高通量测序数据前处理流程。ST 技术常用的前处理工具

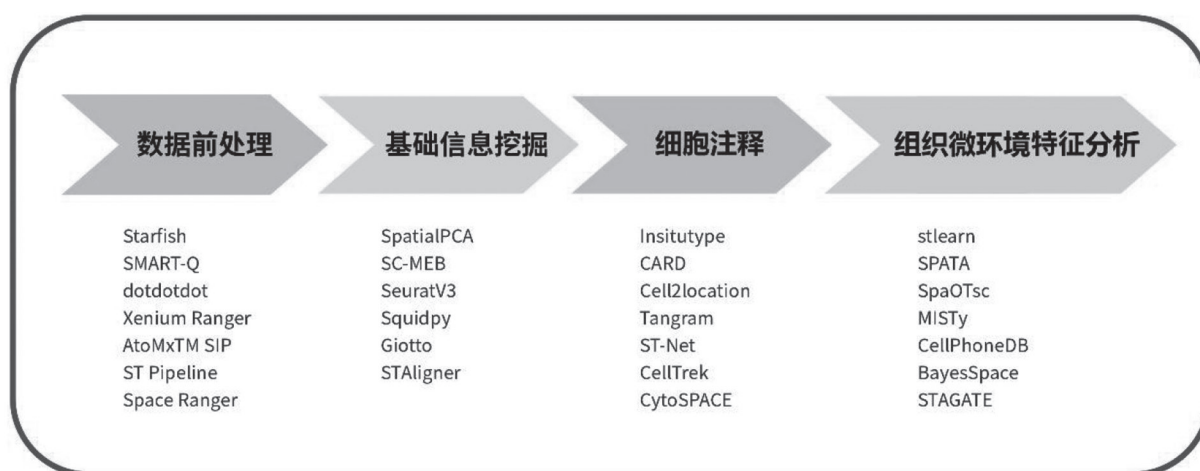


图2 空间转录组数据分析流程图

ST Pipeline^[50] 和 10X Visium 数据专用分析工具 Space Ranger, 都使用 STAR (spliced transcripts alignment to a reference)^[51] 进行序列的拼接与比对工作。

与传统高通量测序数据前处理不同的是, 空间转录组测序结果中还包含了图像数据, 在前处理过程中需要图像与检测区域的配准步骤。而针对 LCM、DSP 等技术获得的测序数据, 前处理步骤与 bulk RNA 测序类似, 且不需要进行额外的图像配准。

2.2 基础生物学信息挖掘

在数据前处理后, 为了从基因表达数据中挖掘具有生物学意义的信息, 需要进行一系列分析, 包括质量控制 (quality control, QC)、数据标准化、筛选高可变基因、降维和聚类、差异表达分析以及个性化分析等。在 QC 步骤, 可以对分割出的“细胞”或 spots 中的 reads 数、基因数等信息进行统计和可视化, 从而把握数据的整体分布特征。在基因表达数据标准化后, 针对数据集的大小筛选高可变基因和降维, 这是因为空间转录组结果中可能包含大量细胞或 spots 的数据, 每个“细胞”或 spot 中又包含了大量基因的表达信息, 会占用大量的时间和分析资源。接下来, 对细胞或 spots 进行聚类分析, 从而将“细胞”或 spots 分配到不同的 cluster 中, 便于探究异质性和差异表达分析。在聚类算法的选择上, 除了传统的 Louvain^[52]、K-means^[53] 等基于基因表达数据进行无监督聚类外, 还有针对空间转录组数据特点开发的聚类算法, 如 SpatialPCA (spatial probabilistic principal component analysis)^[54]、SC-MEB (spatial clustering using the hidden Markov random field based on Empirical Bayes)^[55] 等。聚类后, 便可参考传统转录组差异分析策略计算每个 cluster

的特异高表达基因和富集功能, 确认每个 cluster 的特异性功能 marker。目前, 有很多算法整合了这些基础分析步骤, 使得分析流程一体化, 比如 SeuratV3^[56]、Squidpy (spatial quantification of molecular data in Python)^[57] 和 Giotto^[58] 等等。针对基于空间 barcode 序列的空间转录组技术产生的数据集, 比如 ST、Visium、Stereo-seq 和 Slide-seq 等, 可以使用 STAligner^[59]。STAligner 是一种神经网络模型, 通过生成批量校正的点嵌入来集成和对齐 ST 数据集, 从而实现空间感知数据集集成, 同时进行空间域识别以及下游比较分析。该方法采用图注意力自动编码器学习具有基因表达和空间位置信息的点嵌入, 并构建跨切片的点三元组来指导嵌入对齐过程。该方法也可用于校正批次差异以及 3D 组织结构重建的高级分析。

在空间转录组学研究中, 识别空间可变基因有助于理解疾病的时空特征和组织结构, 其中常用到的算法有 trendsceek^[60]、SpatialDE^[61] 和 STAMarker^[62] 等。

2.3 细胞注释

细胞在空间组织上的分布特征是空间转录组研究的重要方向。因此, 在分析空间转录组数据时需要进行细胞类型的注释。对于 smFISH 和原位测序技术, 在原始数据前处理时已经利用图像分析算法进行细胞分割, 可直接根据每个分割出的“细胞”中的基因表达模式, 参考单细胞转录组细胞注释策略^[63], 基于先验知识或使用自动注释算法如 Insitupype 进行细胞注释。该注释算法基于似然模型, 权衡每个表达值, 提取每个细胞表达谱中的所有可用信息用于注释工作^[64]。而对于以 spot、ROI 为单位进行 bulk 测序的技术, 由于单个区域内的细胞数可能有几个到几百甚至几千个, 包含的细胞类型复

杂多样, 需要新的算法策略进行细胞注释分析。这些算法的主流思路是把单细胞转录组与空间转录组的数据进行联合, 将带有细胞注释信息的单细胞转录组基因表达谱作为参考, 通过解卷积、映射等策略在空间转录组结果中注释细胞类型。解卷积的思路是基于基因表达模式将 spots/ROI 中的多种细胞类型进行拆分, 通过输入单细胞转录组和空间转录组的基因表达信息, 利用不同的算法模型进行分析, 输出每个 spot/ROI 中各细胞类型的比例和丰度。研究人员用 50 个真实和模拟的数据集测试 18 种解卷积算法的准确性、稳健性和可用性, 发现 CARD (conditional autoregressive-based deconvolution)^[65]、Cell2location^[66] 和 Tangram^[67] 的性能表现更好^[68]。

与解卷积不同, 映射的思路是将单细胞转录组和空间转录组, 或空间转录组与病理图像的结果整合到同一维度下进行建模分析, 并基于基因表达数据、病理图像的染色特征和空间转录组数据的坐标信息, 预测细胞 / 基因表达的空间分布模式。ST-Net 是一种深度学习算法^[69], 它结合了空间转录组数据和组织学图像来捕获高分辨率基因表达异质性。从空间转录组 spots 对应位置的组织染色切片中提取出 224×224 像素的斑块区域后, 使用卷积神经网络 DenseNet 模型进行训练和预测, 从而预测平均表达较高的一定数量的基因在组织图像上的表达情况。Wei 等^[70] 开发了 CellTrek 算法, 通过共嵌入方法将单细胞转录组和空间转录组数据整合进一个共享的特征空间, 并使用空间转录组数据训练多元随机森林模型, 结合共享降维特征预测空间坐标。利用训练出的模型计算一个随机森林距离矩阵, 并基于该矩阵预测单细胞转录组数据中细胞的空间坐标。Vahid 等^[71] 开发出 CytoSPACE (cellular spatial positioning analysis via constrained expression alignment) 算法, 这是将单个细胞从单细胞转录组测序图谱映射到空间表达谱的优化方法。该方法将单细胞 /spot 的分配描述为一个凸优化问题, 并使用 Jonker-Volgenant 最短增广路径算法解决该问题, 保证了优秀的映射效果, 抗噪性强, 有助于破译不同生理和病理环境中的单细胞空间变异和群落结构。

上述分析工具的空间聚类和细胞注释示例结果见图 3。

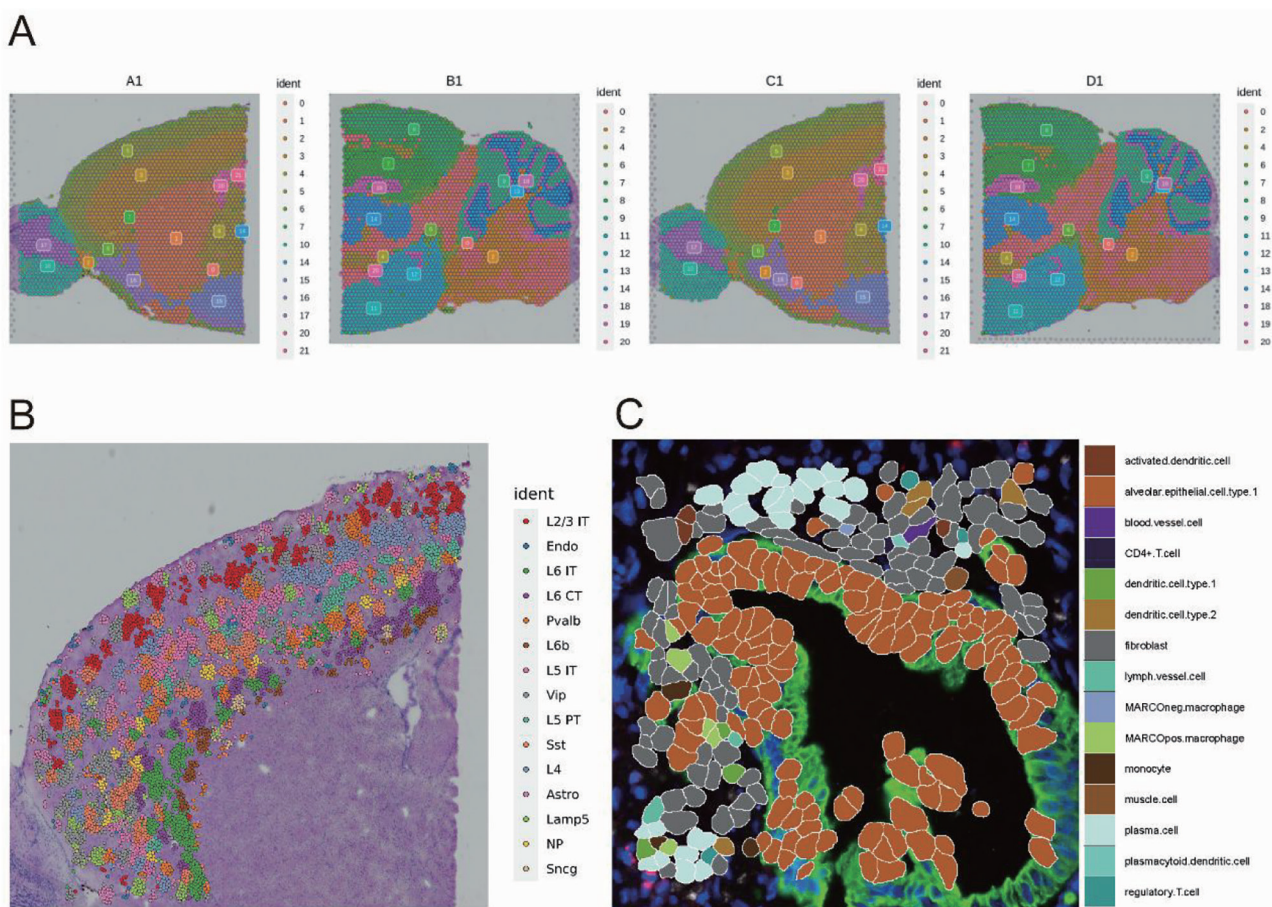
2.4 组织微环境特征分析

组织微环境, 包括细胞外基质 (extracellular matrix, ECM) 和邻近细胞, 是组织形态和功能的基

础, 对组织生长和发育的调控至关重要。当微环境的稳态失衡时, 组织可能会发生炎症反应, 甚至病理性改变。因此, 研究微环境中细胞的轨迹变化、细胞间通讯以及确认组织的空间异质性特征, 是研究疾病发生和发展机制的有效手段。

轨迹分析早在单细胞测序研究中被用于解析细胞的分化轨迹, 并获得基因表达水平随细胞发育轨迹的变化规律。在空间转录组中, 由于一个 spot 或 ROI 可能是多细胞混合的结果, 且具有空间位置信息, 因此需要开发适配空间转录组数据类型的轨迹算法。Pham 等^[72] 开发出的 stLearn 整合了空间聚类、轨迹分析和细胞通讯等分析功能。在 stLearn 中, 空间轨迹分析使用的是自主研发的伪时空轨迹分析 (pseudo-time-space, PSTS) 算法, 可以根据转录组图谱和组织内细胞的空间背景重建进化轨迹。Kueckelhaus 等^[73] 开发出的 SPATA (SPAtial transcriptomic analysis) 需要使用者先在空间转录组结果中预设研究轨迹, 然后分析特定基因或基因集的表达水平沿着轨迹的变化规律, 当轨迹跨越多个微结构时, 使用该方法即可获得基因或基因集在不同组织区域的表达变化轨迹。

空间细胞通讯分析, 顾名思义, 是在空间水平研究不同细胞类型间互作关系的方法。在传统单细胞转录组细胞互作分析的基础上, 还要考虑细胞类型在空间组织的分布情况。Cang 和 Nie^[74] 开发的 SpaOTsc (spatially optimal transporting the single cells) 算法能在单细胞转录组和空间转录组数据间建立映射, 计算细胞间的“空间距离”, 然后重建细胞 - 细胞通讯网络和细胞间的基因 - 基因调控关系。MISTy (multiview intercellular SPAtial modeling framework) 算法是用机器学习对完整的基因表达谱和相互作用进行建模, 建立一种可解释的多视图框架, 而不是分析单个 marker 基因的空间表达模式。该方法允许灵活构建模型来分析空间数据, 而且不需要注释空间单元 (细胞或 spot) 的细胞类型、状态或任何其他特征^[75]。Garcia-Alonso 等^[76] 将单细胞测序中常用的细胞互作算法 CellPhoneDB 升级到了 3.0 版本, 将空间细胞共定位信息加入到配体受体互作分析中。Pham 等^[72] 针对空间转录组的特殊性, 在 stLearn 中开发了 SCTP (spatially-constrained two-level permutation) 算法, 与传统单细胞测序细胞互作分析算法相比, 考虑了基因表达在组织切片上的距离影响。由于细胞间的配体受体互作在 $200 \mu\text{m}$ 距离内结合强度最高, SCTP 分析结果会比传统



(A) Seurat分析工具的多切片空间整合聚类结果。(B) CellTrek细胞映射结果。(C) Giotto分析工具可视化结果, 使用Insituptype进行细胞注释。

图3 空间转录组分析结果图

互作算法具有更低的假阳性率, 从而提高预测准确性。该方法有两种策略, 一种结合邻域分析, 找到共表达的互作关系对显著富集的邻域, 另一种是找到组织中具有较强互作关系的细胞类型组合。使用该方法, 可筛选出组织切片中显著共表达的互作关系对及其空间分布特征, 在空间水平分析细胞通讯网络及其对疾病发展的影响。

空间邻域聚类分析是研究异质性的空间特异性策略, 在分析细胞 /spot 功能时会考虑自身与邻近区域的相互关系, 兼顾本身基因表达模式与空间位置分布。其原理是以某一个细胞 /spot 作为中心, 向其周围扩展一定范围, 将扩展范围内涵盖的细胞 /spots 的数据进行处理以识别空间域, 从而优化聚类结果。使用这种分析策略来进行空间组学分析, 算法有 BayesSpace^[77]、STAGATE^[78] 等。由于这种分析策略使用了空间分布信息, 可以缓解空间聚类结果不连续的问题, 帮助研究者更好地识别组织切

片中不同的亚结构区域, 并解析各空间域的功能异质性。

上述数据分析流程用到的部分空间转录组分析工具见表 2。

3 展望

空间转录组技术的问世改变了生物学基础研究的思路, 能够为科研工作者提供基因表达和细胞类型的空间分布信息, 拓展了组织空间异质性的研究深度。空间技术为研究表观基因组、染色体可及性、染色体结构、蛋白质组、基因治疗、AAV 血清型优化和 TCR 测序带来了新的研究方法。

目前的空间转录组无论在实验技术还是数据分析方法上仍面临诸多挑战和局限性。在实验技术上, 原位检测的分辨率和准确性需要进一步提高, 检测靶标数需要不断增加, 因为高分辨率和覆盖率对于理解肿瘤微环境和转移等至关重要。如何检测非编

表2 空间转录组分析工具列表

算法工具	编程语言	分析步骤	用途	网址	参考文献
Starfish	Python	原始数据前处理	处理基于图像的空间转录组学图像, 对任何FISH方法(smFISH或组合条形码)结果中的RNA转录本进行定位和定量。	https://github.com/spacex/starfish	[47]
SMART-Q	Python	原始数据前处理	一种模块化、可定制的smFISH数据处理流程。同时分析免疫染色信号和 FISH 信号以实现细胞类型特异性的基因定量分析。	https://github.com/shenlab-ucsf/SMART-Q	[48]
dotdotdot	MATLAB	原始数据前处理	一种自动化RNAscope数据定量分析工具, 能够克服组织自发荧光干扰, 多重能力强。	https://github.com/LieberInstitute/dotdotdot	[49]
Xenium Ranger	/	原始数据前处理	提供对 Xenium 原位数据的仪器外再分析。重新标记转录本, 使用最新的分割算法重新分割细胞, 或导入自定义分割数据以将转录本分配给细胞。	https://www.10xgenomics.com/support/software/xenium-ranger/latest	/
Atomx™ SIP	/	原始数据前处理	一种不需要编程经验即可分析CosMx SMI技术产生的数据, 并提供数据可视化功能的云平台, 允许创建和执行自定义分析模块。	https://nanosttring.com/products/atomx-spatial-informatics-platform/atomx-sip-overview/	/
ST Pipeline	Python	原始数据前处理	一种用于分析ST、10X visium技术原始数据的工具, 适合分析带有空间条形码和UMI标识的测序数据。该工具可将基因表达数据定位回其在组织中的位置, 生成用于下游分析的空间转录组数据集。	https://github.com/SpatialTranscriptomics-Research/st_pipeline	[50]
Space Ranger	/	原始数据前处理	10X visium 专用数据分析软件, 可使用明场和荧光显微镜图像处理 Visium 空间转录组数据, 实现基因定量、聚类分析和表达差异分析等目的。	https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/what-is-space-ranger	/
SpatialPCA	R	基础生物学信息挖掘	一种空间感知降维方法, 旨在用低维度展示空间基因表达数据。该方法建立在PCA的基础上, 将定位信息作为附加输入, 并使用核矩阵对组织位置之间的空间相关结构进行显式建模。	https://www.xzlab.org/software.html	[54]
SC-MEB	R	基础生物学信息挖掘	一种使用隐马尔可夫随机场进行空间聚类分析的经验贝叶斯方法, 计算效率高且可扩展至大样本量, 还能够选择平滑参数和聚类数量。	https://github.com/Shufeyangyi2015310117/SC-MEB	[55]
Seurat V3	R	基础生物学信息挖掘	一种整合了多种算法的工具, 能够对单细胞转录组和空间转录组数据集进行质控、探索和整合, 包括高可变基因鉴定、降维聚类、批次效应校正、差异基因筛选及可视化等, 适用于10X Visium、Slide-seq、10X Xenium和Nanostring CosMx SMI等技术数据。	https://satijalab.org/seurat/	[56]

表2 空间转录组分析工具列表(续表)

算法工具	编程语言	分析步骤	用途	网址	参考文献
Squidpy	Python	基础生物学信息挖掘	一种用于空间分子数据分析和可视化的工具, 继承了scanny和anndata的模块化和可扩展性, 汇集了组学和图像分析的工具, 以实现空间转录组数据的可扩展描述, 适用于FISH、10X Visium、Slide-seq、10X Xenium和Nanostring CosMx SMI等技术数据。	https://github.com/scverse/squidpy	[57]
Giotto	R和Python	基础生物学信息挖掘	用于空间数据分析和可视化的工具, 通过各种算法来表征组织组成、空间表达模式和细胞相互作用, 从而提供端到端分析。该工具可以整合单细胞转录组数据进行空间细胞类型富集分析。该方法适用于不同形式的空间转录组数据, 包括seqFISH、MERFISH、STARmap、Visium等。	https://giotto-suite.readthedocs.io/en/master/#	[58]
STAligner	Python	基础生物学信息挖掘	一种神经网络模型, 通过生成批量校正的点嵌入来集成和对齐ST数据集, 从而实现空间感知数据集集成, 同时进行空间域识别以及下游比较分析。主要针对ST、Visium、Stereo-seq和Slide-seq等。	https://github.com/zhoux85/STAligner	[59]
InSituType	R	细胞注释	该算法由Nanostring开发, 基于似然模型, 可以通过贝叶斯分类器从参考数据集进行监督细胞分型, 并通过期望最大化算法进行无监督和半监督细胞分型, 适用于数百万个细胞的数据集, 如Nanostring CosMx SMI。	https://github.com/Nanostring-Biostats/InSituType	[64]
CARD	R	细胞注释	该算法能够适应细胞类型组成中跨组织位置的空间相关性, 实现对细胞类型的解卷积和空间信息的构建, 精准性高。其依赖于一种有效的优化算法进行约束最大似然估计, 可分析具有数万个空间位置和数万个基因的空间转录组数据集。	https://github.com/YMa-lab/CARD	[65]
Cell2location	Python	细胞注释	一个基于贝叶斯模型的解卷积算法, 可以解析空间转录组数据中的细粒度细胞类型。利用负二项式回归整合不同技术和批次的数据, 稳健性好。	https://github.com/BayraktarLab/cell2location/	[66]
Tangram	Python	细胞注释	基于Scanny, 利用深度学习将单细胞(或单核)基因表达数据映射到空间基因表达数据, 可以通过一组共享基因扩展到全基因组范围, 并进行解卷积分析, 或者通过对齐多模态数据以单细胞分辨率解析染色质可及性的空间模式。	https://github.com/broadinstitute/Tangram	[67]
ST-Net	Python	细胞注释	一种深度学习算法, 它结合了空间转录组数据和组织学图像来捕获高分辨率基因表达异质性。该方法适用于ST、Visium技术获得的空间转录组数据集。	https://github.com/bryanhe/ST-Net	[69]

表2 空间转录组分析工具列表(续表)

算法工具	编程语言	分析步骤	用途	网址	参考文献
CellTrek	R	细胞注释	整合scRNA-seq和ST数据, 直接将单个细胞映射回组织切片中的空间坐标。该方法将空间转录组的数据作为训练数据集, 使用随机森林模型预测细胞的空间坐标。	https://github.com/navinlabcode/CellTrek	[70]
CytoSPACE	Python	细胞注释	将单细胞/spot分配制定为凸优化问题, 并使用 Jonker-Volgenant 最短增广路径算法解决该问题, 把单细胞转录组中单个细胞映射到空间转录组数据集中精确的空间位置, 噪声耐受性好。	https://github.com/digitalcytometry/cytospace	[71]
stlearn	Python	组织微环境特征分析	一种将基因表达与空间位置和(或)组织形态信息相结合的工具有效利用空间转录组数据所有的维度。该工具可用于空间聚类、构建时空轨迹(PSTS)、分析空间细胞相互作用(SCTP)等。	https://github.com/Biomedical-MachineLearning	[72]
SPATA	R	组织微环境特征分析	该工具提供空间基因表达的全面表征、转录程序的区域适应和沿空间的瞬态动力学轨迹。	https://github.com/theMILoLab/SPATA	[73]
SpaOTsc	Python	组织微环境特征分析	该工具使用空间转录组数据在单细胞转录组结果中为每个细胞构建空间度量, 从而生成细胞和基因的空间可视化和聚类, 并重建细胞间通讯网络, 推断特定信号传导的空间范围并量化细胞间的基因-基因调控信息流。	https://github.com/zcang/SpaOTsc	[74]
MISTY	R	组织微环境特征分析	一种灵活、可扩展、可解释的机器学习框架, 用于知识提取和空间数据分析, 通过分析细胞内和细胞间的关系, 促进对基因标志物相互作用的深入理解。该框架构建了多个视图以捕捉功能关系, 例如通路活动和交互作用、细胞类型特定关系, 或关注不同解剖区域之间的关系。	https://saezlab.github.io/mistyR/	[75]
CellPhoneDB	Python	组织微环境特征分析	一个公开的人受体、配体互作数据库, 并配备分析细胞间相互作用算法, 3.0版本整合了细胞空间坐标, 可进行空间细胞通讯分析。	https://github.com/Ventolab/CellphoneDB	[76]
BayesSpace	R	组织微环境特征分析	一种完全贝叶斯统计方法, 使用来自空间邻域的信息来增强空间转录组数据的分辨率, 并用于聚类分析。	https://github.com/edward130603/BayesSpace	[77]
STAGATE	Python	组织微环境特征分析	一种整合空间信息和基因表达谱来学习低维潜在嵌入, 从而准确地识别空间域的框架, 采用了一种注意力机制来自适应地学习相邻spot的相似性, 并通过整合基因表达的预聚类来实现可选的细胞类型感知模块。该工具可以扩展到连续切片, 以减少切片之间的批量效应, 并有效地提取三维表达域。	https://github.com/zhanglabtools/STAGATE	[78]

码基因、蛋白质、表观修饰也是技术开发者们亟须解决的问题, 这将有助于更全面地揭示基因调控网络和分子机制。空间转录组技术在单位面积上能够获得丰富的测序数据, 然而, 由于现有数据处理能力的局限, 这一技术的应用仍面临样本尺寸的约束。未来随着技术更新和计算能力的增强, 有望在更大的样本尺寸中检测和分析。

在数据分析上, 需要开发更优质的算法, 从而实现细胞分割更精准, 以及与其他组学数据整合以实现空间单细胞多组学研究的目的是。算法的普及性也仍有待提高, 研究人员仍然经常需要学习新的语法, 甚至新的语言来使用一些数据分析工具, 并且研究者当前高度关注人类和小鼠, 而其他物种的空间转录组学研究滞后。

此外, 如何将空间转录组技术在基础研究中的成果转化到实际的临床应用, 是科研工作者和临床医生面临的重要课题。这包括提高诊断的精度、制定个性化治疗方案以及改善预后管理等。实现这些目标将需要跨学科的合作和持续的技术创新, 为精准医学的发展提供坚实的基础。

空间转录组技术已经展现出强大的应用潜力, 随着实验技术和分析算法的不断发展, 这项技术无疑将引领分子生物学研究的新时代。

[参 考 文 献]

- [1] Tang F, Barbacioru C, Wang Y, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, 2009, 6: 377-82
- [2] Method of the Year 2020: spatially resolved transcriptomics. *Nat Methods*, 2021, 18: 1
- [3] Gall JG, Pardue ML. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci U S A*, 1969, 63: 378-83
- [4] Rudkin GT, Stollar BD. High resolution detection of DNA-RNA hybrids *in situ* by indirect immunofluorescence. *Nature*, 1977, 265: 472-3
- [5] Femino AM, Fay FS, Fogarty K, et al. Visualization of single RNA transcripts *in situ*. *Science*, 1998, 280: 585-90
- [6] Lubeck E, Cai L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods*, 2012, 9: 743-8
- [7] Lubeck E, Coskun AF, Zhiyentayev T, et al. Single-cell *in situ* RNA profiling by sequential hybridization. *Nat Methods*, 2014, 11: 360-1
- [8] Chen KH, Boettiger AN, Moffitt JR, et al. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, 2015, 348: aaa6090
- [9] Chang CA, Ahle D, Urdea MS, et al. Characterization of synthetic branched DNA (BDNA) using polyacrylamide gel-filled capillary electrophoresis. *Nucleic Acids Symp Ser*, 1991, (24): 223
- [10] Player AN, Shen LP, Kenny D, et al. Single-copy gene detection using branched DNA (bDNA) *in situ* hybridization. *J Histochem Cytochem*, 2001, 49: 603-12
- [11] Battich N, Stoeger T, Pelkmans L. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat Methods*, 2013, 10: 1127-33
- [12] Wang F, Flanagan J, Su N, et al. RNAscope: a novel *in situ* RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn*, 2012, 14: 22-9
- [13] Baner J, Nilsson M, Mendel-Hartvig M, et al. Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res*, 1998, 26: 5073-8
- [14] Dirks RM, Pierce NA. Triggered amplification by hybridization chain reaction. *Proc Natl Acad Sci U S A*, 2004, 101: 15275-8
- [15] Lee JH, Daugharthy ER, Scheiman J, et al. Highly multiplexed subcellular RNA sequencing *in situ*. *Science*, 2014, 343: 1360-3
- [16] Lee JH, Daugharthy ER, Scheiman J, et al. Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc*, 2015, 10: 442-58
- [17] Ke R, Mignardi M, Pacureanu A, et al. *In situ* sequencing for RNA analysis in preserved tissue and cells. *Nat Methods*, 2013, 10: 857-60
- [18] Bustin S, Dhillon HS, Kirvell S, et al. Variability of the reverse transcription step: practical implications. *Clin Chem*, 2015, 61: 202-12
- [19] Alon S, Goodwin DR, Sinha A, et al. Expansion sequencing: spatially precise *in situ* transcriptomics in intact biological systems. *Science*, 2021, 371: eaax2656
- [20] Wang X, Allen WE, Wright MA, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, 2018, 361: eaat5691
- [21] Stahl PL, Salmen F, Vickovic S, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 2016, 353: 78-82
- [22] Rodrigues SG, Stickels RR, Goeva A, et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 2019, 363: 1463-7
- [23] Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 2015, 161: 1202-14
- [24] Russell A, Weir JA, Nadaf NM, et al. Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. *Nature*, 2024, 625: 101-9
- [25] Vickovic S, Eraslan G, Salmen F, et al. High-definition spatial transcriptomics for *in situ* tissue profiling. *Nat Methods*, 2019, 16: 987-90
- [26] Liu Y, Yang M, Deng Y, et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell*, 2020, 183: 1665-81
- [27] Chen A, Liao S, Cheng M, et al. Spatiotemporal

- transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, 2022, 185: 1777-92
- [28] Fu X, Sun L, Dong R, et al. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell*, 2022, 185: 4621-33
- [29] Meier-Ruge W, Bielser W, Remy E, et al. The laser in the Lowry technique for microdissection of freeze-dried tissue slices. *Histochem J*, 1976, 8: 387-401
- [30] Emmert-Buck MR, Bonner RF, Smith PD, et al. Laser capture microdissection. *Science*, 1996, 274: 998-1001
- [31] Zhu YY, Machleder EM, Chenchik A, et al. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques*, 2001, 30: 892-7
- [32] Hashimshony T, Wagner F, Sher N, et al. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*, 2012, 2: 666-73
- [33] Ramskold D, Luo S, Wang YC, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol*, 2012, 30: 777-82
- [34] Junker JP, Noel ES, Guryev V, et al. Genome-wide RNA Tomography in the zebrafish embryo. *Cell*, 2014, 159: 662-75
- [35] Okamura-Oho Y, Shimokawa K, Takemoto S, et al. Transcriptome tomography for brain analysis in the web-accessible anatomical space. *PLoS One*, 2012, 7: e45373
- [36] Combs PA, Eisen MB. Sequencing mRNA from cryo-sliced *Drosophila* embryos to determine genome-wide spatial patterns of gene expression. *PLoS One*, 2013, 8: e71820.
- [37] Lovatt D, Ruble BK, Lee J, et al. Transcriptome *in vivo* analysis (TIVA) of spatially defined single cells in live tissue. *Nat Methods*, 2014, 11: 190-6
- [38] Merritt CR, Ong GT, Church SE, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol*, 2020, 38: 586-99
- [39] Lin PB, Tsai AP, Soni D, et al. INPP5D deficiency attenuates amyloid pathology in a mouse model of Alzheimer's disease. *Alzheimers Dement*, 2023, 19: 2528-37
- [40] Xie L, Kong H, Yu J, et al. Spatial transcriptomics reveals heterogeneity of histological subtypes between lepidic and acinar lung adenocarcinoma. *Clin Transl Med*, 2024, 14: e1573
- [41] Delorey TM, Ziegler C, Heimberg G, et al. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature*, 2021, 595: 107-13
- [42] Janesick A, Shelansky R, Gottscho AD, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and *in situ* analysis. *Nat Commun*, 2023, 14: 8353
- [43] He S, Bhatt R, Brown C, et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol*, 2022, 40: 1794-806
- [44] Jiang F, Zhou X, Qian Y, et al. Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nat Methods*, 2023, 20: 1048-57
- [45] Sudmeier LJ, Hoang KB, Nduom EK, et al. Distinct phenotypic states and spatial distribution of CD8⁺ T cell clonotypes in human brain metastases. *Cell Rep Med*, 2022, 3: 100620
- [46] Bressan D, Battistoni G, Hannon GJ. The dawn of spatial omics. *Science*, 2023, 381: eabq4964
- [47] Perkel JM. Starfish enterprise: finding RNA patterns in single cells. *Nature*, 2019, 572: 549-51
- [48] Yang X, Bergenholtz S, Maliskova L, et al. SMART-Q: an integrative pipeline quantifying cell type-specific RNA transcription. *PLoS One*, 2020, 15: e228760
- [49] Maynard KR, Tippani M, Takahashi Y, et al. dotdotdot: an automated approach to quantify multiplex single molecule fluorescent *in situ* hybridization (smFISH) images in complex tissues. *Nucleic Acids Res*, 2020, 48: e66
- [50] Navarro JF, Sjostrand J, Salmen F, et al. ST Pipeline: an automated pipeline for spatial mapping of unique transcripts. *Bioinformatics*, 2017, 33: 2591-3
- [51] Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 2013, 29: 15-21
- [52] Subelj L, Bajec M. Unfolding communities in large complex networks: combining defensive and offensive label propagation for core extraction. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2011, 83: 36103
- [53] Hartigan JA, Wong MA. Algorithm AS 136: a K-means clustering algorithm. *J R Stat Soc C*, 1979, 28: 100-8
- [54] Shang L, Zhou X. Spatially aware dimension reduction for spatial transcriptomics. *Nat Commun*, 2022, 13: 7203
- [55] Yang Y, Shi X, Liu W, et al. SC-MEB: spatial clustering with hidden Markov random field using empirical Bayes. *Brief Bioinform*, 2022, 23: bbab466
- [56] Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. *Cell*, 2019, 177: 1888-902
- [57] Palla G, Spitzer H, Klein M, et al. Squidpy: a scalable framework for spatial omics analysis. *Nat Methods*, 2022, 19: 171-8
- [58] Dries R, Zhu Q, Dong R, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol*, 2021, 22: 78
- [59] Zhou X, Dong K, Zhang S. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nat Comput Sci*, 2023, 3: 894-906
- [60] Edsgard D, Johnsson P, Sandberg R. Identification of spatial expression trends in single-cell gene expression data. *Nat Methods*, 2018, 15: 339-42
- [61] Williams CG, Lee HJ, Asatsuma T, et al. An introduction to spatial transcriptomics for biomedical research. *Genome Med*, 2022, 14: 68
- [62] Zhang C, Dong K, Aihara K, et al. STAMarker: determining spatial domain-specific variable genes with saliency maps in deep learning. *Nucleic Acids Res*, 2023, 51: e103
- [63] Clarke ZA, Andrews TS, Atif J, et al. Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat Protoc*, 2021, 16:

- 2749-64
- [64] Danaher P, Zhao E, Yang Z, et al. Insitutype: likelihood-based cell typing for single cell spatial transcriptomics. *bioRxiv*, 2022, <https://doi.org/10.1101/2022.10.19.512902>
- [65] Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat Biotechnol*, 2022, 40: 1349-59
- [66] Kleshchevnikov V, Shmatko A, Dann E, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol*, 2022, 40: 661-71
- [67] Biancalani T, Scalia G, Buffoni L, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods*, 2021, 18: 1352-62
- [68] Li H, Zhou J, Li Z, et al. A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nat Commun*, 2023, 14: 1548
- [69] He B, Bergenstrahle L, Stenbeck L, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng*, 2020, 4: 827-34
- [70] Wei R, He S, Bai S, et al. Spatial charting of single-cell transcriptomes in tissues. *Nat Biotechnol*, 2022, 40: 1190-9
- [71] Vahid MR, Brown EL, Steen CB, et al. High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. *Nat Biotechnol*, 2023, 41: 1543-8
- [72] Pham D, Tan X, Balderson B, et al. Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. *Nat Commun*, 2023, 14: 7739
- [73] Kueckelhaus J, von Ehr J, Ravi VM, et al. Inferring spatially transient gene expression pattern from spatial transcriptomic studies. *bioRxiv*, 2020, <https://doi.org/10.1101/2020.10.20.346544>
- [74] Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat Commun*, 2020, 11: 2084
- [75] Tanevski J, Flores R, Gabor A, et al. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biol*, 2022, 23: 97
- [76] Garcia-Alonso L, Handfield LF, Roberts K, et al. Mapping the temporal and spatial dynamics of the human endometrium *in vivo* and *in vitro*. *Nat Genet*, 2021, 53: 1698-711
- [77] Zhao E, Stone MR, Ren X, et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol*, 2021, 39: 1375-84
- [78] Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun*, 2022, 13: 1739