



空间转录组的解读与探究: 数据分析方法学研究现状与展望

李润泽^{†*}, 陈旭[†], 杨雪瑞^{*}

生物信息学教育部重点实验室, 合成与系统生物学中心, 清华大学生命科学学院, 北京 100084

[†] 同等贡献

* 联系人, E-mail: lirz2023@mail.tsinghua.edu.cn; yangxuerui@tsinghua.edu.cn

2024-03-28 收稿, 2024-05-29 修回, 2024-07-08 接受, 2024-07-09 网络版发表

国家重点研发计划(2023YFC3043300)、国家自然科学基金(32330022, 81972912)和清华大学自主科研计划资助

摘要 作为近年来快速发展的新兴技术, 空间转录组学已经极大地改变了生物和医学多个领域的研究范式. 这一技术保留了复杂组织中细胞的空间定位信息, 并且以多细胞团、单细胞或亚细胞分辨率进行转录组分析. 细胞空间定位信息与其分子特征谱的耦合产生了新型的多模态高通量数据资源, 这对高效的数据挖掘分析方法的开发提出了新的挑战. 由于生理组织样本的复杂性与技术的局限性, 空间转录组数据具有高度的非理想性. 其数据结构复杂、信噪比低、稀疏性强且覆盖度不均一, 给数据的深度分析和生物学信息的解析带来了一系列挑战, 目前许多困难尚未解决. 此外, 多个技术路线的空间转录组学正在快速迭代发展, 需要现有数据分析理论的发展和创新技术工具的开发. 本文概述了当前空间转录组数据的常用解析方法, 针对不同角度与层面的信息挖掘需求, 讨论了现有方法的设计思路与潜在的局限性, 并对未来方法学设计的方向和策略进行了展望, 以期激发对新的数据分析理论、模型和算法的讨论和开发.

关键词 空间转录组, 生物信息学, 人工智能, 机器学习, 数据挖掘

生理和病理组织在细胞类型和细胞的空间排列方面表现出多重异质性^[1]. 单细胞RNA测序(single cell RNA sequencing, scRNA-seq)技术通过揭示单个细胞的分子表达谱, 推动了对组织异质性的理解, 已经有大量研究利用这一技术阐明细胞多样性和组织/器官的空间结构^[1~4]. 尽管scRNA-seq数据适用于解读细胞异质性^[5~7], 但由于空间信息的丢失, 它难以用于研究细胞内分子谱与原生组织微环境之间的相互影响.

随着高通量测序、微阵列、并行高分辨成像、显微切割与捕获等技术的进步, 在保留组织中细胞空间信息的前提下, 定量分析其转录组特征的技术正在快速出现和发展^[8], 并推进至单细胞、亚细胞解析度, 极

大地扩展了我们对神经系统^[9]、胚胎发育^[10]、肿瘤^[11]等多种复杂生理与病理组织的认识. 总体上, 目前的空间转录组分析技术主要有两类策略: 基于高通量测序和基于成像, 它们能够以不同的分辨率和尺度进行(亚)转录组分析. 其中, 空间转录组学(spatial transcriptomics, ST)^[9,12]、Slide-seq^[13]、用于空间组学测序的组织确定性条形码(deterministic barcoding in tissue for spatial omics sequencing, DBiT-seq)^[14]、Seq-Scope^[15]、空间分辨转录本扩增子读出映射(spatially-resolved transcript amplicon readout mapping, STARmap)^[16]、扩张测序(expansion sequencing, Exseq)^[17]、聚合酶链反应索引文库测序(polony-indexed library-sequencing,

引用格式: 李润泽, 陈旭, 杨雪瑞. 空间转录组的解读与探究: 数据分析方法学研究现状与展望. 科学通报, 2024, 69: 4415–4431

Li R Z, Chen X, Yang X R. Decoding spatial transcriptomics: Current trends and future prospects in data analysis methodologies (in Chinese). Chin Sci Bull, 2024, 69: 4415–4431, doi: [10.1360/TB-2024-0332](https://doi.org/10.1360/TB-2024-0332)

Pixel-seq)^[18]、多组学原位成对测序(multi-omics *in situ* pairwise sequencing, MIP-seq)^[19], 以及空间增强分辨率组学测序(spatial enhanced resolution omics-sequencing, Stereo-seq)^[10]等基于高通量测序的方法解决了先前空间编码和原位测序规模上的限制. 顺序荧光原位杂交(sequential fluorescence *in situ* hybridization, seq-FISH)^[20]和多重误差稳健杂交(multiplexed error-robust FISH, MERFISH)^[21]等基于成像的技术利用高分辨显微镜, 通过荧光原位杂交产生多重分子图谱. 与此同时, 不同路线的空间转录组实验分析技术仍在快速迭代发展, 以进一步提高检测分辨率、吞吐量、灵敏度和鲁棒性等^[22–24].

空间转录组学为我们理解复杂组织中的生理功能提供了丰富的信息. 首先, 空间转录组学可以深入了解细胞相互作用的空间特征, 这在传统的scRNA-seq检测中难以实现^[25–27]. 此外, 空间转录组学能够阐明组织内的空间基因表达模式, 从而揭示单个细胞亚群的空间特化^[28]. 更重要地, 空间转录组学是在空间与分子维度上揭示各种疾病模式的强大工具^[29,30]. 例如, 肿瘤异质性不仅体现在一系列不同细胞群的特征, 还表现在各种细胞亚群之间的局部聚集和相互作用, 而这些信息都被编码在空间转录组图谱中^[31,32]. 因此, 肿瘤空间转录组数据为了解驱动恶性肿瘤的细胞亚群提供了重要的见解, 有助于开发针对肿瘤特定活动区域的靶向疗法^[33]. 最后, 空间转录组学有助于探索发育生物学机制, 例如胚胎发生^[10,34]和神经发生过程^[35–37]. 相比之下, scRNA-seq可以阐明细胞如何进化成不同的细胞类型, 而空间转录组学则揭示了各种细胞类型在器官发生过程中如何组织的空间线索.

总之, 空间转录组学填补了基于传统分离测序或低通量成像方法的关键空白, 代表了多细胞系统原位解码的技术前沿. 然而, 空间数据具有高度复杂性(多维性、多模态和非线性)和非理想性(高噪声、低且不均匀的覆盖率以及高丢失率)^[38,39], 这需要方法理论创新和数据深度分析技术的进一步发展. 本文剖析了空间转录组学面临的分析任务, 介绍了现有的解决方案, 讨论了其基本原理和假设, 并进一步讨论了其适应性与潜在的局限性. 空间转录组数据的分析工作流程通常由几个阶段组成^[40], 其中包括数据预处理和分析、生物学见解的挖掘和空间维度下的知识重建(图1). 在每个阶段, 我们将探索当前使用的计算技术和模型, 并确定迫切需要改进的领域. 具体而言, 本文将介绍空间

反卷积(图1(a))、基因插补(图1(b))、图像对齐和细胞分割(图1(c))等数据预处理方法, 以及空间基因表达模式的鉴定、空间数据的整合(图1(d))、空间聚类(图1(e))、基于scRNA-seq数据的空间重建(图1(f))、细胞互作的推断(图1(g))等分析方法. 此外, 我们还将深入讨论空间数据的模拟方法, 这是一个在近一年内迅速发展但并未在以前的综述中涉及的新领域. 我们希望这篇综述能够为利用空间转录组学进行生物学发现的研究人员提供最新的参考.

1 空间转录组数据的预处理

空间转录组数据的处理涉及一系列精细的步骤, 旨在优化数据质量, 提高分析的精确度和可靠性. 预处理、空间反卷积和空间插补这三个部分共同构成了这一流程的核心, 反映了一个共性目标: 通过各种策略最大化原始数据中的生物学信息提取, 同时最小化技术噪音的影响. 在这个过程中, 从原始数据的预处理到复杂数据的反卷积和插补, 每一步都需要精心设计的算法和方法, 应对不同的技术挑战, 从而确保数据的准确解读, 保证数据转化过程的科学性和有效性.

鉴于空间转录组数据的高度复杂性, 正确的预处理对于后续分析和信息挖掘至关重要. 具体的预处理策略需要根据特定的空间转录组技术进行调整, 其最终目标是在减少系统性偏差并规范数据格式的同时, 保留必要的空间信息.

对于基于成像的空间转录组技术, 初始图像的预处理包括图像校正、拼接、配准和分割^[40], 这一阶段的主要目的是消除技术伪影、提高信噪比, 并实现读数标准化. 具体而言, 图像配准可以对齐多个z轴上的堆栈或连续切片, 纠正在实验过程中产生的组织结构偏移^[41,42]. 基于FISH的技术能够在单细胞甚至亚细胞水平上获得高分辨率数据, 但也因此在空间的数据预处理方面出现了额外的挑战. 例如, 细胞分割任务需要识别细胞边界和其他细胞结构. 然而, 由于细胞密度和复杂的细胞形状等因素, 自动细胞分割是一项复杂的任务^[43,44]. 近年来不断开发的分割方法已经逐渐提高了分割的准确性和质量^[45–48].

深度学习的最新进展使细胞分割和图像配准自动化任务取得了显著的进步^[49–53]. 由于深层神经网络能够有效地从复杂数据中学习复杂模式和表示, 其在细胞分割的尺寸识别、边界平滑以及鲁棒性方面都表现出了较传统方法更出色的性能^[52,53]. 最近的STaCker集

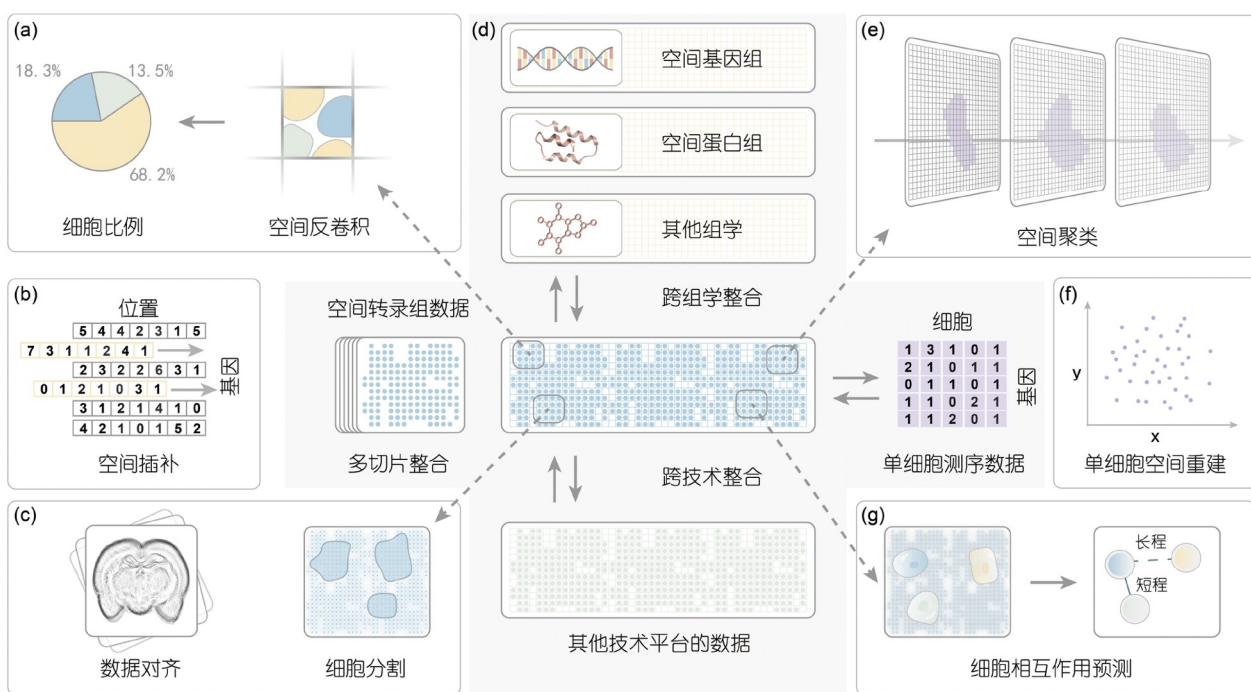


图 1 空间转录组数据的分析任务概述。(a) 空间反卷积任务。该分析通过使用scRNA-seq数据作为参考,对捕获区域单元内的转录组谱进行反卷积,推断区域单元内细胞种类及其表达特征。(b) 基因插补任务。该分析用于推断特定组织位置内缺失的基因表达数据。(c) 图像对齐和细胞分割任务。(d) 空间数据的整合分析,包括连续多切片的整合(左)、空间多组学的整合(上)、和scRNA-seq数据的集成(右)以及跨技术平台的异构数据整合(下)。(e) 空间聚类,包括识别2D平面和由连续切片构成的3D空间中的组织学特征域。(f) scRNA-seq数据的空间重建任务。(g) 基于ST数据推断细胞间相互作用。其中包括长程和短程互作等不同模式

Figure 1 Overview of spatial transcriptome data analysis tasks. (a) Spatial deconvolution task. This analysis uses scRNA-seq data as a reference to deconvolute the transcriptome profiles within captured spatial units, inferring the cell types and their expression characteristics within these units. (b) Gene imputation task. This analysis is used to infer missing data of gene expression at specific tissue locations. (c) Image alignment and cell segmentation task. (d) Integration analysis of spatial data. It includes the integration of consecutive slices (left), spatial multi-omics integration (top), integration with scRNA-seq data (right), and heterogeneous data integration across different technological platforms (bottom). (e) Spatial clustering. This involves identifying histological signature domains in both 2D planes and 3D spaces constructed from consecutive slices. (f) Spatial reconstruction of scRNA-seq data. (g) Inference of cell-cell interactions based on ST data. This includes different modes of long-range and short-range interactions

成组织图像和基因表达为复合图像,使用包含多个二维卷积层的U-net网络对齐切片,对数据噪声和批次效应表现出了较之前方法更高的鲁棒性^[54]。

对于在空间网格内捕获转录本的非成像空间转录组学技术,预处理步骤较成像技术具有很大的差异。这些步骤需要将捕获的转录本映射到原本的空间坐标^[55,56]。此外,空间坐标映射与之匹配的组织图像能够生成空间基因表达的整体视图。

在确保数据质量的基础上,空间转录组数据处理的一个重要环节是通过空间反卷积从数据中识别和量化不同细胞类型的贡献。具体而言,空间转录组技术直接采集的一个空间点可能包含来自多个细胞的转录本,为了提高识别细胞亚群的精度和组织图谱的准确性,空间反卷积算法被用于推断每个捕获位置内不同细胞

类型的相对比例。目前用于空间反卷积的工具包括基于参考的方法和无参考的方法,具体取决于当前空间转录组数据是否存在匹配的scRNA-seq数据集^[57]。

基于参考的方法使用配对的scRNA-seq数据中细胞类型的表达谱对空间位置内的细胞组成进行推断。各种不同的策略被用于整合scRNA-seq和空间转录组数据。例如,Slide-seq和SPOTlight使用非负矩阵分解回归(NMF)重建细胞类型的加权组合^[13,58]。SpatialDWLS采用两步策略,包括细胞类型富集和细胞类型反卷积的加权最小二乘推理^[59]。由于不同技术平台产生的数据集往往表现出细胞类型的异质性,RCTD采用了考虑平台效应的最大似然估计方法,有效增强了其跨平台准确性的表现^[60]。另外,贝叶斯推理方法(例如cell2location和DestVI)和基于概率的建模方法(Steroscope)在

细胞类型反卷积任务上表现良好^[61~63]。采用非凸优化方法并整合成像数据的Tangram能够对多种类型空间转录组数据进行反卷积^[64]。

无参考方法不需要额外数据的介入, 相较于基于参考的方法而言具有更好的通用性。一些无参考方法可以利用来自同一组织的公开scRNA-seq数据来减少技术噪音, 并实现跨样本和平台的数据整合。例如, CARD是一种对邻域相似性进行建模的条件自回归反卷积方法, 在scRNA-seq参考不匹配的情况下也能够进行细胞类型反卷积^[65]。最近的SpSeudoMap方法基于排序的scRNA-seq数据生成细胞混合物^[66], 以此对空间转录组数据进行模仿以产生类似的基因表达模式, 随后通过CellDART对细胞类型进行预测并调整scRNA-seq数据以纠正参考不匹配的问题^[67]。

更进一步地, 一些方法可以在没有参考的情况下执行空间反卷积, 整个流程仅依赖于空间组学数据。这类方法可以是数据驱动的, 例如STdeconvolve^[68]使用潜在狄利克雷分配来推断细胞类型, 也可以是由标记基因指导的, 例如Cellscope^[69]、ST-Assign^[70]和SMART^[71], 其中标记基因是来自公共数据库或文献的先验。具体而言, ST-Assign基于标记基因使用Metropolis-within-Gibbs采样器同时进行细胞类型注释和分解^[70], 而SMART方法使用半监督模型同时推断细胞类型组成和细胞类型-特定基因表达谱^[71]。

在空间转录组学数据处理中的另一个关键步骤是空间基因的插补, 其目标是使用计算方法来推断特定组织位置缺失的基因表达数据。类似MERFISH的高分辨率空间转录组学技术可能无法捕获组织内所有基因的表达信息, 插补任务的目的是填补这些缺失值以获得更完整的基因表达谱。类似于反卷积任务, 大多数基因插补工具也需要使用scRNA-seq数据参考。例如, GimVI是一种基于分层贝叶斯模型的概率方法, 通过非线性变换生成基因分布参数以集成scRNA-seq数据^[72]。另一种常用的策略是使用线性方法生成嵌入表示, 其中一个例子是Seurat, 它基于典型相关分析(CCA)进行数据降维, 并根据共表达模式估计基因表达^[73]。Liger和Harmony是类似的线性嵌入方法^[74,75], 它们都将scRNA-seq和ST数据映射到相同的空间以进行基因插补。另外, SpaGE是一种基于k近邻的预测方法, 表现出比上述方法更好的性能^[76]。StPlus基于自动编码器, 相较仅使用共享基因的方法, 它利用scRNA-seq数据中包含的整体信息来获取更全面的参考^[77]。除了能

够完成空间反卷积任务, Tangram还可以进行基因插补, 它通过最小化特征之间的余弦距离来建立从scRNA-seq到ST的映射矩阵^[64]。另外, TransImp使用正则化来减轻对空间自相关的高估, 并能够提供估算的不确定性分数以评估可靠性^[78]。与上述方法不同的是, 最新的工具STAGE^[79]不需要利用scRNA-seq数据作为参考, 它基于空间监督的自编码器结构来学习空间表达模式, 并为给定的空间坐标生成高分辨率的表达谱, 克服了稀疏空间分辨率和低数据质量的限制。

2 组织内特征区域的鉴定

复杂组织内的异质性体现在由不同分子与细胞特征所定义的不同区域模块。这些特征区域可能由同种细胞或不同种类细胞组成, 其内部具有特定的基因表达模式或多种分子特征的关联, 并与细胞间组装、组织的形成与生理功能紧密相关^[80]。因此, 基于空间聚类来识别复杂组织中的特征区域已成为近年来空间转录组数据的关键分析。

传统的细胞聚类方法依赖基因表达水平, 但在应用于空间转录组数据时, 可能会错误地将具有相似表达谱的不同空间区域分为一组, 导致组织区域内聚类结果不连续。这可能会歪曲真实的空间模式并阻碍潜在功能相关域的识别, 进一步混淆对空间邻域细胞互作的生物学见解。为了解决这一限制, 学术界开发了一系列计算方法, 整合空间、分子特征以及细胞形态信息进行细胞聚类及空间分区, 在多种组织中鉴定了具有潜在生理学功能特征的空间区域^[81]。这些方法包括基于特征提取和常规聚类的策略^[82], 基于马尔可夫随机场的策略^[28,83], 基于空间统计学的方法^[84], 以及基于图深度学习的算法^[85,86]。

首先, 基于特征提取和常规聚类的方法借助预训练模型(例如: CNN)从HE图像中提取组织学特征, 随后利用Kmeans进行聚类^[87]。其次, 通过比较空间点与其周边邻居的相似性, 隐马尔科夫随机场HMRF(hidden Markov random field)模型鼓励相邻像素点具有相同的聚类标签。这种方法考虑了周边细胞基因表达模式的联合概率, 有利于发现连续的空间组织模式。另外, HMRF方法可以准确地检测和解剖结构相符的空间特征域^[28]。基于空间统计的策略主要是受到地理学空间分析方法的启发, 它们利用空间统计测试来识别空间域(例如, Moran's I^[88]、高斯过程回归、空间点回归等)。最后, 与以上相对传统的方法相比, 基于图卷积网

络等图深度学习的模型能够从更加整体和深层的角度合并空间特征,推断数据像素点的潜在结构,并且更好地克服空间转录组数据的复杂性和高噪音,产生更加细化连续且有生理学意义的聚类谱^[85,86,89-91]。与前两种方法相比,图卷积网络和其他图深度学习方法从更全面和深层的角度整合空间特征。它们通过捕获空间像素点的潜在结构,表现出更加细化连续的聚类谱^[85,86,89-91]。

通过鉴定特征区域和空间上连贯的细胞群,空间聚类方法为我们提供了深入了解细胞组织模式和相互作用的独特视角。随着领域内对多切片和多技术数据集分析的关注不断增加,对空间域识别性能的要求也将随之提升。此外,现有方法识别的空间域通常过于离散且彼此独立,未来的方法应侧重于揭示不同区域之间的功能交互。这需要对聚类结果有更加深入的生物学解释,现有的方法正在启发进一步的验证与分析,以了解空间聚类模式是否定义了新的解剖区域特征,以及所识别的新细胞簇是否具有功能性的意义^[25]。

3 空间基因表达异质性的解析

识别空间差异基因(SVG)是一项关键分析任务,用于揭示组织内部的空间表达模式。大多数早期方法基于统计测试,通常遵循“两步走”策略:如图2所示,首先将稀疏空间网格的表达值建模为空间坐标的回归函数;随后,基于该回归模型,基因表达的变化被分解为空间协方差和独立噪声,其中空间协方差反映了基因的表达水平是否依赖于空间分布,表明基因表达的不同空间模式(图2)。早期统计方法的差异主要在于对数据分布的假设及对丢失值、技术噪声和随机方差等数据特征的建模方式。

例如, SpatialDE^[92]和 Trendsceek^[93]作为两种早期的空间差异基因鉴定方法,分别利用了高斯过程回归和标记点过程模型。它们都对原始计数数据进行了分布转换。与此不同的是, SPARK^[94]则直接对原始计数值进行建模,并使用广义线性模型中的泊松分布等各种内核来获得标准校正的 P 值。这使得SPARK能够更好地控制 I 类错误并获得更高的检测灵敏度。然而,早期方法在面对大规模数据集时会产生巨大的计算成本。为了解决计算复杂性的问题, SPARK-X^[95]提出了一种可扩展的非参数模型,它使用代数运算构建基因表达和空间坐标的协方差矩阵。类似地, BOOST-GP^[96]同样直接对原始计数值建模,但采用了负二项式(negative

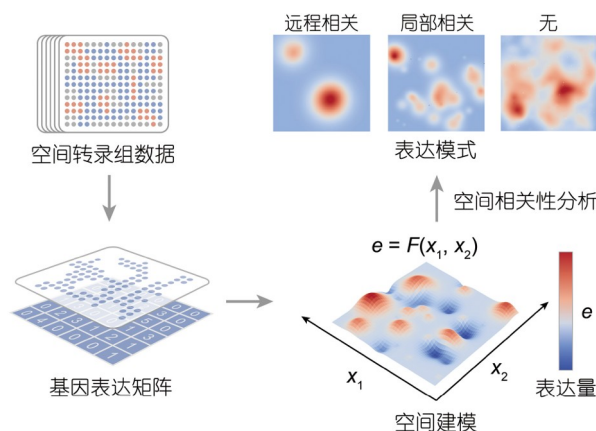


图2 基于统计方法识别空间差异基因的流程示意图。在空间转录组数据中,二维切片空间中每个基因的表达值可以表示为二维矩阵。经典的统计模型将基因表达视为空间坐标 x_1 和 x_2 的回归函数,通过拟合获得回归且平滑的表达分布。随后,基于该回归模型,基因表达的变异被分解为空间协方差和独立噪声。空间协方差分量反映了不同基因表达水平相对于空间位置的相关性,用于捕获不同的空间模式(例如:远程相关、局部相关和不受空间位置影响)

Figure 2 Schematic diagram of identifying spatially variable genes using statistical methods. In spatial transcriptome data, the expression value of each gene in one slice can be represented as a 2D matrix. Classical statistical models regard gene expression as a regression function of spatial coordinates x_1 and x_2 , obtaining a regressed and smoothed expression distribution through fitting. Subsequently, based on this regression model, the variance in gene expression is decomposed into spatial covariance and independent noise. The spatial covariance component reflects the correlation of different gene expression levels relative to spatial locations, capturing various spatial patterns (e.g., long-range correlation, local correlation, and patterns unaffected by spatial location)

binomial, NB)分布,这一假设非常适合解释测序数据中经常观察到的过度分散现象,与广泛使用的RNA-seq分析工具相一致,例如DESeq2^[97]和edgeR^[98]。除了数据分布的建模之外,不同的方法也在数据噪声和内在方差的基本假设上有显著区别。例如, NoVaTeST^[99]解决了现有方法经常忽视的一个关键方面,即评估基因表达的噪声方差是否依赖于位置或异方差,进而识别异方差噪声的程度,其中表达噪声的方差随空间坐标而变化。除了上述方法中使用的空间统计模型外,一些方法还采用隐马尔可夫随机场(HMRF)模型,例如scGCO^[100]。HMRF模型捕获变量的统计(条件独立性)和空间特性,为识别SVG提供独特的视角。简而言之,这些统计方法仍然保留了在协方差中添加空间成分以识别显著空间变异基因的核心原则。

统计建模通常依赖于固定的参数假设,这些假设可能不足以捕捉复杂的空间关系。基于计算密集型的重采样测试也受到诸多限制。为了解决这些问题,基于

图的深度学习方法在识别SVG方面受到了广泛的关注。SOMDE^[101]使用自组织映射神经网络和高斯过程对空间数据进行建模,在保证SVG基因识别质量的前提下维持较高的计算效率。SpaGCN^[85]是一种图卷积网络方法,它集成了基因表达数据、空间位置信息和组织学图像。GLISS^[102]同样是一种基于图的方法,不同的是,它将细胞位置视为连续变量,更准确地捕获空间表达梯度上的连续性。然而,这些方法大多数不容易识别空间域特异性的可变基因,限制了某些特定的下游分析。为了解决这个问题,STAMarker^[103]采用由图注意力自编码器、多层感知分类器和显著性图组成的三阶段框架,可用于识别特定于空间域的SVG基因。

总之,空间差异基因的识别是理解复杂空间模式的一项关键任务。基于该任务提供的信息,我们可以对组织内不同区间的基因变化模式进行分类,识别空间热点区域,以帮助我们分析特定表型、功能或生理过程中的细胞组装、迁移等问题,并发现与之相关的关键基因。现有方法仍存在计算开销和统计解析效率的问题。为了平衡这两个问题,可能需要灵活的非参数平滑方法,典型的例子是最近的可扩展工具SMASH^[104]。此外,新兴算法已经证明了利用多模态单细胞空间特征(包括RNA、蛋白质和染色质可及性)能够改进识别效果^[105],这凸显了空间多模态数据集成的必要性。

4 细胞-细胞互作的推断

多细胞组织的生理功能不仅取决于异质性细胞的类型组成,也高度依赖于细胞之间复杂相互作用的动态稳定^[106-108]。因此,了解细胞间相互作用对于组织功能、发育和复杂疾病的研究至关重要。日益积累的scRNA-seq和空间转录组数据,为系统地破译细胞互作及其分子机制提供了更深入的视角。单细胞转录组学的研究已经从仅仅识别组织中的细胞类型,发展到更深入地关注细胞之间的相互关系。这种研究模式的转变促进了大量的计算方法和分析策略的出现,以应对越来越复杂的数据分析需求^[27,109]。

针对传统的scRNA-seq数据,已经有许多计算方法用于推断细胞互作^[5,110-114],或分析与细胞互作相关的细胞内信号通路^[115-120]。这些方法主要依赖于配体-受体对共表达的设计原理,领域内已经存在一系列综述文章对这些方法进行了详细介绍^[27,109,121,122],在此不再赘述。而在空间转录组研究的早期阶段,通过整合空

间转录组和scRNA-seq数据的优势^[123,124],细胞的空间距离信息可用过滤和纠正从scRNA-seq数据推断出的细胞互作关系^[25]。许多计算工具和模型采用了这种设计策略^[27,125,126],其中包括(1)基于区域聚类或基于反卷积的方法和(2)基于最优传输理论的方法。鉴于大多数空间转录组数据并非单细胞尺度,第一类是目前更常用的方法。

进一步地,研究者开发了专门适用于空间转录组数据的计算方法。这些方法无须输入scRNA-seq数据,根据分析场景和设计原理可以分为两类:(1)基于配体-受体对空间共表达的细胞互作推断^[127-132];(2)量化细胞互作对基因表达的效应或识别相互作用的基因^[133-137]。其中,第一类方法采用的模型包括空间自相关模型、差异基因分析、统计检验和矩阵分解等^[129-132]。这些模型擅长揭示局部空间中配体-受体对的共表达模式。相比之下,第二类方法则采用了更精细复杂的模型,包括高斯过程建模^[137]、图神经网络^[138]、集成学习^[133,135]、卷积神经网络^[136]等。其中,高斯过程建模适用于捕捉基因表达中受细胞互作影响的空间变化,但其计算成本高且对核函数的选择敏感^[137];图神经网络通过图结构来模拟细胞相互作用,尽管其参数调整相对复杂;集成学习则结合了多种模型的优势^[136],增强了对不同细胞互作模式的识别能力;卷积神经网络在提取层次特征方面表现优异,但其结果的可解释性存在问题。这些模型擅长捕获局部依赖性并揭示空间共表达的配体-受体对。

综上所述,无论是整合scRNA-seq和空间转录组数据,亦或是直接使用空间转录组数据推断细胞互作,通常都基于细胞互作的先验性信息。然而,此类方法容易受到数据特征的不均衡空间分布的影响,并且强烈依赖配-受体对的先验知识或预定义的特征(如:细胞类型信息)。事实上,细胞互作模式的多样性和复杂性远超配体-受体对所能覆盖的范围^[27,139]。此外,大多数细胞的配-受体表达信息往往残缺不全,这使得全面识别细胞互作关系变得十分困难。更重要的是,这些方法在重构受到严格空间约束的局部相互作用时表现较好,但对于那些不受或少受空间位置约束的长程相互作用,其有效性、准确性和系统性较差,获得的细胞互作结果也常常是动态变化的^[26,140,141]。以上问题强调了发展新的理论假设和方法的必要性和紧迫性。为了应对这些挑战,我们开发了DeepLinc工具^[142],它基于假设邻近细胞之间的多模式相互作用比随机选择的远距离细胞

更为丰富,通过变分图自动编码器框架,从头重建完整的细胞互作网络.该方法为理解细胞间复杂的互作关系提供了新的思路和工具.

5 空间转录组数据的模拟

长期以来,空间转录组数据分析领域缺乏用于基准测试和算法验证的标准数据集.现有的scRNA-seq模拟流程由于无法自然扩展至空间转录组数据,在面临额外空间信息时显得力不从心.空间转录组数据模拟方法的出现,使这一问题得到了部分解决,更好地支持分析方法的开发和验证.此类方法旨在重现真实且多样化的空间转录组数据,同时允许对算法和模型进行质控和评估^[57,143].

在空间转录组数据模拟的早期阶段,研究者们通常采用简单的随机生成策略,例如点交换、随机重采样或伪点创建等^[85,92,144–147].这些方法对基因表达数据的分布进行了特定假设^[145],因此在再现真实组织的复杂空间模式方面存在一定的局限性.这些方法对基因表达数据的分布进行了特定假设^[145],使得它们在复现真实组织的复杂空间模式方面受到一定的限制.另一方面,由于额外空间信息的存在,现有的scRNA-seq模拟流程无法自然扩展到空间转录组数据.因此,需要新的、精细设计的模拟方法来生成真实且多样化的空间数据.为了应对这些挑战,近年来一些研究者结合生物学知识、实验技术差异和先进的统计建模技术,开发了更复杂的空间转录组数据模拟方法^[143,148–151].例如,SRtsim^[148]是第一个跨技术进行空间解析转录组数据模拟的计算框架,如10× Visium^[9,12]、MERFISH^[21]和seqFISH+^[152].具体而言,SRtsim使用凹包算法生成位置坐标,包括方形网格或随机点过程形式,并且支持无参考模拟,允许用户从头定义和创建不同的空间表达模式.scDesign3^[150]则采用统一的概率模型模拟单细胞和空间组学数据,能够推断具有生物学意义的参数,并评估生成的细胞簇、轨迹和空间位置的拟合优度.这种统一的方法简化了不同实验场景的模拟过程,确保在多个组学中一致且高质量地生成合成数据.

然而,SRtsim和scDesign3主要模拟以点或空间像素单元的粗粒度数据,未能充分捕捉组织的精细化结构特征,例如脉管系统、组织区室和细胞外基质等.最近的工具spaSim^[151]通过硬核过程或均匀间隔模型来创建细胞背景,并采用随机数采样来分配肿瘤组织和

正常组织中的细胞类型,从而捕捉了特定区域的细胞浸润/排斥以及血液/淋巴管的特征,产生比基于点模式更具代表性和真实性的空间模拟^[148,150].类似地,Spider^[149]和synthspot^[143]也采用了在定制化的空间区域内采样和分配不同类型细胞的设计理念.不同之处在于,Spider利用启发式的批量退火算法,使其能够在五分钟内生成100万个细胞的空间转录组数据;而synthspot依赖于均匀分布的随机采样,这需要更多时间并生成粗颗粒度数据.SRtsim和scDesign3主要模拟以点或空间像素单元的粗粒度数据.然而,更真实的模拟应该捕捉组织的精细化结构特征,例如脉管系统、组织区室和细胞外基质等,这些特征会影响细胞的空间分布和基因表达.最近的工具spaSim^[151]通过硬核过程或均匀间隔模型来创建细胞背景,并且采用随机数采样来分配肿瘤组织和正常组织中的细胞类型.这一设计捕获了基本的空间特性,例如特定区域的细胞浸润/排斥以及血液/淋巴管的特征,从而产生比基于点模式更具代表性和真实性的空间模拟^[148,150].类似地,Spider^[149]和synthspot^[143]也采用了在定制化的空间区域内采样和分配不同类型细胞的设计理念.不同之处在于,Spider利用启发式的批量退火算法,使其能够在5 min内生成100万个细胞的空间转录组数据;相比之下,synthspot依赖于均匀分布的随机采样,这需要更多时间并生成粗颗粒度数据.

总之,空间转录组数据的模拟是一个快速发展的新兴领域,主要目的是提供更严格的模拟算法和方法评估工具.当前的方法主要集中在静态数据、多细胞尺度和单一模态等方面.未来的研究方向应更加关注模拟方法的真实性和复杂性,尤其是在模拟微环境和真实组织的动态特征方面.为了实现这一目标,除了考虑各种空间平台的固有噪声和系统差异,还应结合时间维度的信息,捕捉空间异质性随时间变化的动态过程.这不仅要求从亚细胞到组织水平的全面模拟,还需涵盖从转录组到多分子特征的多尺度、多模拟数据.通过结合实际组织的内在性质,合成数据集可以更准确地反映真实生物系统的复杂性,从而为测试和验证分析方法提供更可靠的参考.此外,开发用户友好型的交互式模拟平台将是一个重要的方向,这些平台应允许研究人员直观地设置和观察各种参数下的模拟结果,提高模拟过程的可操作性和可访问性,通过这些平台,研究人员可以更方便地进行实验设计和数据分析,推动空间转录组学研究的进一步发展.

6 空间数据的深度整合

单切片的空间转录组数据仅提供组织内细胞的二维快照, 尽管有时配对的scRNA-seq数据可作为参考, 但在捕捉复杂的空间生物学时, 尤其是涉及三维结构和细胞互动的情境下, 其完整性和准确性往往受限。因此, 连续切片的对齐和整合对重建真实的三维组织结构至关重要。

为了解决这一问题, 研究者们开发了基于概率统计和深度学习策略^[153,154]的多种整合分析工具。例如, PASTE采用Gromov-Wasserstein最优传输策略来对齐和融合不同切片之间的信息。尽管这种方法提高了细胞类型和空间可变基因的识别效率, 但却显著增加了计算成本。与PASTE不同, GraphST^[153]结合了图神经网络和自监督对比学习, 通过数据驱动的方法来对齐连续的组织切片, 减少了批次效应并提高了空间反卷积和聚类任务的性能。此外, STitch3D^[155]和Spatial-Prompt^[156]也通过不同的架构(如图注意力网络和非负岭回归)进行连续切片的整合, 进一步推进了空间反卷积和空间域的识别。最新的工具SPACEL^[157]采用图卷积网络和对抗性学习算法, 能够识别跨多个切片的空间相关域, 并且基于统一的空间坐标系重建组织的三维结构。除了图网络之外, 领域对抗神经网络也被用于多切片集成, 例如DeeST^[154]工具。

除了单技术平台的数据整合, 多源数据的整合也日益重要。不同技术平台的数据在实验条件、批次和技术上存在显著差异, 这对数据匹配和整合提出了更高的要求。为了解决这一问题, 研究者们开发了结合互近邻算法(mutual nearest neighbors, MNN)和图深度学习技术的方法^[158-160]。例如, STAligner^[158]通过图注意力自编码器和三元组学习策略有效消除批次效应; SPIRAL采用了GraphSAGE^[161]结构和域适应技术, 优化了不同批次坐标的对齐; SLAT使用图对抗匹配策略, 通过建模细胞关系来进行空间数据的对齐。

空间多模态数据的整合分析也备受关注。这些数据包括组织学图像、基因表达谱和细胞空间分布等多种类型, 它们的整合有助于全面解析复杂生物系统的结构和功能。例如, ST-Net^[162]从苏木精和伊红染色的组织病理学图像中预测局部基因表达。这类方法直接从组织图像预测空间分辨转录组, 有利于基于图像筛选具有空间异质性的分子标志物。ST-Net建立了基因表达和组织形态之间的定量联系, 进一步地,

SpaCell^[163]基于自编码器结构, 通过整合成像数据和基因表达识别细胞类型和预测疾病进展阶段。stLearn^[164]基于统计建模, 通过整合图像、基因表达和空间分布, 以更稳定和更高的准确性进行空间轨迹推断、细胞互动分析和空间插补。最新的工具STASCAN^[165]作为一种新颖的深度学习方法, 通过结合基因表达谱和组织学图像的细胞特征来定义捕获区域和未知区域的细胞类型分布。与其他方法不同的是, STASCAN揭示了更精细的组织结构, 并重建了更详细的3D细胞模型。这意味着STASCAN大大提高了空间细胞分辨率, 在利用组织学图像生成细胞图谱方面具有显著的优势。

此外, 空间多组学数据的融合也是一个快速发展的前沿领域, 研究者们开发了多种分析方法, 结合不同的分子信息解析复杂的空间生物学特征。其中包括空间转录组和其他空间组学数据(例如空间蛋白质组学和空间基因组学)的集成。受到GLUE^[166]算法的启发, SpatialGlue^[167]结合空间蛋白质组学和空间基因组学数据, 通过先验的引导图将不同组学特征嵌入到共同的低维空间中, 基于特征重要性的评价分数, 获得了不同的生物学发现。除了空间生物学注释外, 空间多组学整合也有利于空间可变基因的进一步挖掘。例如, 最近推出的spatial IMIX^[168]框架整合了空间多组学数据, 进一步降低了空间差异基因的假阳性率。这种跨模态的整合可以提供更全面的生物学视角, 揭示转录水平与各种生物过程(例如蛋白质表达和基因组变异)之间复杂的相互作用。事实上, 多批次数据合并的方法通常优于一般性的下游分析策略。因此, 开发有效的数据集成策略不仅是推进空间转录组研究的先决条件, 也是深入阐明生物学问题的关键一步。

7 观点与展望

空间转录组学技术为理解组织功能和发育过程提供了丰富的信息。然而, 这些组织切片的快照无法全面地展现组织原位信息, 尤其是细胞间的相互作用。同时, 由于技术与科学原理的限制, 大多数单细胞空间组学数据具有非线性、噪音水平高、缺失值多、高稀疏性以及数据覆盖度较低且不均衡等方面的不足。近年来, 学术界开发了一系列算法工具, 深入解读数据的潜在信息, 挖掘复杂组织的区域化功能特征(图1)。相应地, 最近还发展了一系列的大规模空间组学数据库和交互式的通用分析平台^[169-175], 它们进一步推动了计算方法的发展, 促进了空间组织结构的深度解析。尽管如此,

现有方法对于信息挖掘的广度与深度仍非常有限,亟待分析方法学理论的突破,适应快速发展的实验技术。未来的方法学研究需要解决的潜在挑战讨论如下:

7.1 数据复杂度和特征维度的增加

随着空间转录组分析技术的快速发展,数据的复杂性和维度也在不断增加。新一代的分析方法应当有效地解决以下问题,挖掘更加全面和深入的信息(图3)。

(1) 三维空间的数据分析,具体表现为跨连续切片^[176,177]或多个分层^[136](图3(a))。

(2) 时序数据分析,用于捕获正常或病理组织随时间变化的空间动态^[10,178](图3(b))。

(3) 空间多组学平台和技术产生的多模态数据格式,包括图像、序列和数据点阵等^[179-181](图3(c),(d))。

(4) 空间分辨率的提高,表现为从多细胞到单细胞甚至亚细胞水平(图3(e))。

(5) 分子特征的插补和扩展。

(i) 三维数据的扩展。与单一切片的数据相比,连续多个超薄切片可以更好地近似生物组织的真实3D结构。目前大多数连续切片的集成分析方法仍然基于对

齐和拼接的策略^[44,153]。此外,这类方法没有充分利用3D信息,仅执行空间聚类或空间差异基因识别等基本任务,这导致数据的利用程度不够,可能无法捕捉真实组织三维空间中的细微差异和新颖机制。因此,我们认为未来可以探索针对3D切片数据的新分析任务,以扩大切片集成方法的研究范围。另一方面,样品制备、保存和染色方法的改进可能有助于减少伪影并保持生物样品的原始空间结构,这对于3D空间数据的集成和分析至关重要。

(ii) 时间维度的引入。不同时间点所收集的空间组学数据可用于揭示生物系统的时空动态。在时间尺度上分析空间数据具有很大的挑战性,迫切需要先进的技术来解释组织内的时空模式。首先,当前的方法多为离散点时间采样,未来可能致力于提高时间分辨率,同时应当开发方法跟踪分子的连续变化事件。此外,另一个重要方向是发现组织内细胞或分子的时空轨迹,这将揭示空间基因表达的变化和构建动态基因调控网络。

(iii) 多模态和多组学数据的可扩展和可解释性分析框架。组织学图像^[182]和多组学数据^[167]是空间数据典型的多模态数据。尽管多模式数据的整合有利于促

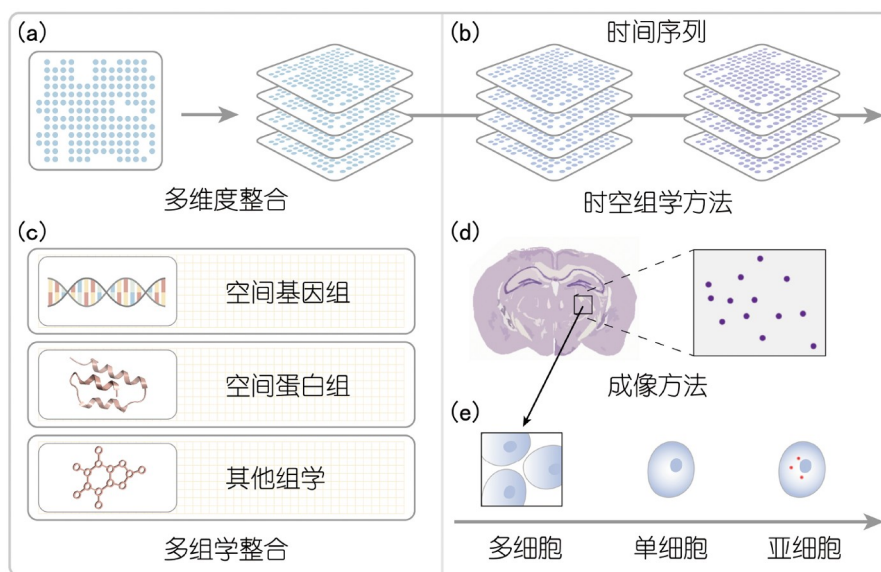


图3 数据复杂性和特征维度的增加。(a) 从2D到3D的维度扩展,意味着有必要对多个连续或非连续切片进行整合和组织重建。(b) 时空组学可以捕获空间分子谱的动态变化。(c) 空间多组学平台的发展,例如空间基因组学和空间蛋白质组学。(d) 空间转录组技术产生的多种数据格式,例如组织学图像和荧光阵列。(e) 空间分辨率从多细胞提高到单细胞甚至亚细胞水平

Figure 3 Increasing data complexity and feature dimensions. (a) Expansion from 2D to 3D Dimensions. This implies the necessity for the integration and reconstruction of multiple consecutive or non-consecutive slices. (b) Spatiotemporal omics can capture the dynamic changes in spatial molecular profiles. (c) Development of spatial multi-omics platforms. For example, spatial genomics and spatial proteomics. (d) Various data formats generated by spatial transcriptomics technologies, such as histological images and fluorescence arrays. (e) Spatial resolution enhancement from multicellular to single-cell and even subcellular levels

进生物学信息的全面挖掘,然而在保留空间背景的前提下,实现整合的过程是相当具有挑战性的.对于异源异构样本,除了消除实验的批次效应,相似样本的链接和锚定也是一个重要的问题.此外,生物学家对不同组学层之间的调控关系尤其感兴趣,但是现有方法无法很好地捕获这些关系.一般而言,这些工具通常采用模型的事后解释策略,基于同一细胞群内的组学相关性来连接各组学特征.实际上,这种非端到端的方法忽略了整体上跨层的相互作用^[167,168],导致空间多组学整合和分析之间存在方法的割裂关系.单细胞RNA-seq数据分析中也存在类似的问题^[166,183,184].我们认为,迁移学习可能是一种潜在的解决方案,因为它可以自然地利用跨组学层的信息流动,从而实现生物知识的迁移而不仅仅是数据层面的信息交换^[185].

(iv) 亚细胞解析度的信息挖掘.随着亚细胞级空间组学技术的快速发展,空间聚类分析将逐渐从细胞水平过渡到亚细胞水平.理论上,高分辨率空间组学数据可用于解析局部空间内分子簇的共定位和聚集模式.然而,由于单一切片数据的高噪声水平和横截面特征,其所捕获的仅是各细胞内分子定位的部分信息,并不能反映分子定位的全貌.为了克服这一限制,应当考虑利用连续多切片的数据分析不同细胞类型中的多分子定位模式.该分析需要两个主要步骤:首先,需要分子在亚细胞结构(例如内质网、高尔基体和细胞膜)上定位的参考信息.建立这样的参考文献和数据库是一个值得发展的重要研究方向;其次,需要精细设计的算法来准确识别共定位的分子簇,进而揭示多样化和复杂的分子定位机制.

(v) 基于空间数据特征设计的分子插补策略.现有大多数空间转录组技术的基因覆盖度非常低,即使对于高通量空间转录组技术,在大多数细胞中往往只检测到几百个转录本,这意味着每个细胞内的分子位置在很大程度上是随机的抽样测定.因此,开发高效插补亚细胞空间数据的方法是未来重要的发展方向.目前,大多数基因插补方法^[72~78]几乎都使用scRNA-seq数据

作为参考,恢复空间转录组数据中的缺失值.实际上,这些方法遵循与常规scRNA-seq数据相同的插补原理^[186,187],然而对于空间数据,更重要的任务是恢复空间中丢失的转录本位置,而不仅仅是表达值.具体而言,分子位置信息的丢失一般体现在组织切片的横轴和纵轴两个维度.首先,一些低分辨率图像重建的方法^[188]可能有助于恢复横轴空间像素中丢失的转录本信息.其次,细胞内分子的分布会在切片纵轴重叠和混合.纵轴信息的丢失可能导致3D空间的亚细胞分析出现重大错误.因此,开发新的方法来恢复细胞内分子沿纵轴的分布是一个潜在的研究方向.这可能会催生新一代的空间分析算法,以扩展空间数据的数据维度和信息.

7.2 知识迁移促进新的生物学发现

知识迁移是将一种背景下的见解应用到另一种背景中的过程.与空间转录组数据的独立分析相比,与其他数据或知识源(例如基因组、蛋白质组、临床数据和分子调控网络等)之间的知识迁移策略更有利于全面解析生物系统的复杂性.具体而言,典型思路是利用注释良好的数据(例如scRNA-seq)来促进空间数据的挖掘.例如,将空间转录组数据中学到的细胞互作模式迁移到大规模常规scRNA-seq数据中是一个重要的任务,这使我们能够预测、探索和验证更多异质性组织中的细胞互作.另一个可能的应用是利用空间转录组和空间蛋白质组之间的共定位特征推断空间上的翻译动态.此外,多切片数据能够提供转录本3D分布模式的分析,有望为全细胞模拟提供参考.最后,整合细胞内分子调控网络等先验知识,可以帮助解释空间组数据中细胞互作模式的生物学相关性.

总而言之,实验和计算方法的协同进步能够充分发挥空间转录组学的潜力.随着技术的更新迭代,我们相信空间组学领域将获得系统性的发展,包括具有更强大、更灵活的分析框架,更好的数据和信息集成策略,更高的预测效率和准确性,以及更深入地考虑数据和生物过程的复杂性.

致谢 衷心感谢国家蛋白质科学中心清华分中心(北京)和清华大学蛋白质研究技术中心的支持,以及清华大学生物计算平台、基因组测序与分析平台等核心设施的支持.

参考文献

- 1 Bich L, Pradeu T, Moreau J F. Understanding multicellularity: The functional organization of the intercellular space. *Front Physiol*, 2019, 10: 1170

- 2 Rao L, Cai L, Huang L. Single-cell dynamics of liver development in postnatal pigs. *Sci Bull*, 2023, 68: 2583–2597
- 3 Wang Y, Zhao Y, Chen S, et al. Single cell atlas of developing mouse dental germs reveals populations of CD24⁺ and Plac8⁺ odontogenic cells. *Sci Bull*, 2022, 67: 1154–1169
- 4 Li Z, Yang Q, Tang X, et al. Single-cell RNA-seq and chromatin accessibility profiling decipher the heterogeneity of mouse $\gamma\delta$ T cells. *Sci Bull*, 2022, 67: 408–426
- 5 Efremova M, Vento-Tormo M, Teichmann S A, et al. CellPhoneDB: Inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc*, 2020, 15: 1484–1506
- 6 Nitzan M, Karaikos N, Friedman N, et al. Gene expression cartography. *Nature*, 2019, 576: 132–137
- 7 Ren X, Zhong G, Zhang Q, et al. Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res*, 2020, 30: 763–778
- 8 Burgess D J. Spatial transcriptomics coming of age. *Nat Rev Genet*, 2019, 20: 317
- 9 Ståhl P L, Salmén F, Vickovic S, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 2016, 353: 78–82
- 10 Chen A, Liao S, Cheng M, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, 2022, 185: 1777–1792.e21
- 11 Moncada R, Barkley D, Wagner F, et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol*, 2020, 38: 333–342
- 12 Rao N, Clark S, Habern O. Bridging genomics and tissue pathology. *Genet Eng Biotechnol News*, 2020, 40: 50–51
- 13 Rodrigues S G, Stickels R R, Goeva A, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 2019, 363: 1463–1467
- 14 Liu Y, Yang M, Deng Y, et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell*, 2020, 183: 1665–1681.e18
- 15 Cho C S, Xi J, Si Y, et al. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell*, 2021, 184: 3559–3572.e22
- 16 Wang X, Allen W E, Wright M A, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, 2018, 361: eaat5691
- 17 Alon S, Goodwin D R, Sinha A, et al. Expansion sequencing: Spatially precise *in situ* transcriptomics in intact biological systems. *Science*, 2021, 371: eaax2656
- 18 Fu X, Sun L, Dong R, et al. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell*, 2022, 185: 4621–4633.e17
- 19 Wu X, Xu W, Deng L, et al. Development of multiomics *in situ* pairwise sequencing (MiP-Seq) for single-cell resolution multidimensional spatial omics. *bioRxiv*, 2023, doi: 10.1101/2023.01.07.523058
- 20 Shah S, Lubeck E, Zhou W, et al. *In situ* transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron*, 2016, 92: 342–357
- 21 Chen K H, Boettiger A N, Moffitt J R, et al. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, 2015, 348: aaa6090
- 22 Moffitt J R, Lundberg E, Heyn H. The emerging landscape of spatial profiling technologies. *Nat Rev Genet*, 2022, 23: 741–759
- 23 Cheng M, Jiang Y, Xu J, et al. Spatially resolved transcriptomics: A comprehensive review of their technological advances, applications, and challenges. *J Genet Genomics*, 2023, 50: 625–640
- 24 Liu B, Li Y, Zhang L. Analysis and visualization of spatial transcriptomic data. *Front Genet*, 2022, 12: 785290
- 25 Walker B L, Cang Z, Ren H, et al. Deciphering tissue structure and function using spatial transcriptomics. *Commun Biol*, 2022, 5: 220
- 26 Liu Z, Sun D, Wang C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biol*, 2022, 23: 218
- 27 Armingol E, Officer A, Harismendy O, et al. Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet*, 2021, 22: 71–88
- 28 Zhu Q, Shah S, Dries R, et al. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence *in situ* hybridization data. *Nat Biotechnol*, 2018, 36: 1183–1190
- 29 Park H, Jo S H, Lee R H, et al. Spatial transcriptomics: Technical aspects of recent developments and their applications in neuroscience and cancer research. *Adv Sci*, 2023, 10: 2206939
- 30 Zheng B, Fang L. Spatially resolved transcriptomics provide a new method for cancer research. *J Exp Clin Canc Res*, 2022, 41: 179
- 31 Puram S V, Tirosh I, Parkh A S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*, 2017, 171: 1611–1624.e24
- 32 Wagner J, Rapsomaniki M A, Chevrier S, et al. A single-cell atlas of the tumor and immune ecosystem of human breast cancer. *Cell*, 2019, 177:

1330–1345.e18

- 33 Wu S Z, Al-Eryani G, Roden D L, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet*, 2021, 53: 1334–1347
- 34 Xu Y, Zhang T, Zhou Q, et al. A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nat Cell Biol*, 2023, 25: 604–615
- 35 Chen X, Sun Y C, Zhan H, et al. High-throughput mapping of long-range neuronal projection using *in situ* sequencing. *Cell*, 2019, 179: 772–786. e19
- 36 Zhang M, Eichhorn S W, Zingg B, et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature*, 2021, 598: 137–143
- 37 Ortiz C, Carlén M, Meletis K. Spatial transcriptomics: Molecular maps of the mammalian brain. *Annu Rev Neurosci*, 2021, 44: 547–562
- 38 Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods*, 2022, 19: 534–546
- 39 Zeng Z, Li Y, Li Y, et al. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biol*, 2022, 23: 83
- 40 Dries R, Chen J, del Rossi N, et al. Advances in spatial transcriptomic data analysis. *Genome Res*, 2021, 31: 1706–1718
- 41 Coskun A F, Cai L. Dense transcript profiling in single cells by image correlation decoding. *Nat Methods*, 2016, 13: 657–660
- 42 Bryan J P, Binan L, McCann C, et al. Optimization-based decoding of imaging spatial transcriptomics data. *Bioinformatics*, 2023, 39: btad362
- 43 Borovec J, Kybic J, Arganda-Carreras I, et al. ANHIR: Automatic non-rigid histological image registration challenge. *IEEE Trans Med Imag*, 2020, 39: 3042–3052
- 44 Zeira R, Land M, Strzalkowski A, et al. Alignment and integration of spatial transcriptomics data. *Nat Methods*, 2022, 19: 567–575
- 45 Thomas R M, John J. A review on cell detection and segmentation in microscopic images. In: Proceedings of the 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT), 20–21 April 2017, 2017
- 46 Vicar T, Balvan J, Jaros J, et al. Cell segmentation methods for label-free contrast microscopy: Review and comprehensive comparison. *BMC BioInf*, 2019, 20: 360
- 47 Petukhov V, Xu R J, Soldatov R A, et al. Cell segmentation in imaging-based spatial transcriptomics. *Nat Biotechnol*, 2022, 40: 345–354
- 48 Littman R, Hemminger Z, Foreman R, et al. Joint cell segmentation and cell type annotation for spatial transcriptomics. *Mol Syst Biol*, 2021, 17: e10108
- 49 Araújo F H D, Silva R R V, Ushizima D M, et al. Deep learning for cell image segmentation and ranking. *Computized Med Imag Graphics*, 2019, 72: 13–21
- 50 Lugagne J B, Lin H, Dunlop M J, et al. DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. *PLoS Comput Biol*, 2020, 16: e1007673
- 51 Greenwald N F, Miller G, Moen E, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol*, 2022, 40: 555–565
- 52 Chen H, Li D, Bar-Joseph Z. SCS: Cell segmentation for high-resolution spatial transcriptomics. *Nat Methods*, 2023, 20: 1237–1243
- 53 Wang Y, Wang W, Liu D, et al. GeneSegNet: A deep learning framework for cell segmentation by integrating gene expression and imaging. *Genome Biol*, 2023, 24: 235
- 54 Peter L, Shawn M, Kun X, et al. Image guided construction of a common coordinate framework for spatial transcriptome data. *bioRxiv*, 2023, 2023.2011.2014.565868
- 55 Zhou Y, Jia E, Pan M, et al. Encoding method of single-cell spatial transcriptomics sequencing. *Int J Biol Sci*, 2020, 16: 2663–2674
- 56 Teves J M, Won K J. Mapping cellular coordinates through advances in spatial transcriptomics technology. *Mol Cells*, 2020, 43: 591–599
- 57 Li B, Zhang W, Guo C, et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat Methods*, 2022, 19: 662–670
- 58 Elosua-Bayes M, Nieto P, Mereu E, et al. SPOTlight: Seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res*, 2021, 49: e50
- 59 Dong R, Yuan G C. SpatialDWLS: Accurate deconvolution of spatial transcriptomic data. *Genome Biol*, 2021, 22: 145
- 60 Cable D M, Murray E, Zou L S, et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotechnol*, 2022, 40: 517–526
- 61 Kleshchevnikov V, Shmatko A, Dann E, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol*, 2022, 40: 661–671
- 62 Lopez R, Li B, Keren-Shaul H, et al. DestVI identifies continuums of cell types in spatial transcriptomics data. *Nat Biotechnol*, 2022, 40: 1360–1369
- 63 Andersson A, Bergenstråhle J, Asp M, et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun Biol*, 2020, 3: 565
- 64 Biancalani T, Scalia G, Buffoni L, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods*, 2021, 18: 1352–1362
- 65 Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat Biotechnol*, 2022, 40: 1349–1359

- 66 Bae S, Choi H, Lee D S. spSeudoMap: Cell type mapping of spatial transcriptomics using unmatched single-cell RNA-seq data. *Genome Med*, 2023, 15: 19
- 67 Bae S, Na K J, Koh J, et al. CellDART: Cell type inference by domain adaptation of single-cell and spatial transcriptomic data. *Nucleic Acids Res*, 2022, 50: e57
- 68 Miller B F, Huang F, Atta L, et al. Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nat Commun*, 2022, 13: 2339
- 69 Geras A, Darvish Shafighi S, Domżał K, et al. Celloscope: A probabilistic model for marker-gene-driven cell type deconvolution in spatial transcriptomics data. *Genome Biol*, 2023, 24: 120
- 70 Agnieszka G, Ewa S. ST-Assign: A probabilistic model for joint cell type identification in spatial transcriptomics and single-cell RNA sequencing data. *bioRxiv*, 2023, 2023.2005.2029.542559
- 71 Yang C, Sin D D, Ng R T. SMART: Reference-free deconvolution for spatial transcriptomics using marker-gene-assisted topic models. *bioRxiv*, 2023, 2023.2006.2020.545793
- 72 Lopez R, Nazaret A, Langevin M, et al. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. *arXiv preprint*, arXiv: 190502269, 2019
- 73 Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. *Cell*, 2019, 177: 1888–1902.e21
- 74 Welch J D, Kozareva V, Ferreira A, et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell*, 2019, 177: 1873–1887.e17
- 75 Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*, 2019, 16: 1289–1296
- 76 Abdelaal T, Mourragui S, Mahfouz A, et al. SpaGE: Spatial gene enhancement using scRNA-seq. *Nucleic Acids Res*, 2020, 48: e107
- 77 Shengquan C, Boheng Z, Xiaoyang C, et al. stPlus: A reference-based method for the accurate enhancement of spatial transcriptomics. *Bioinformatics*, 2021, 37: i299–i307
- 78 Chen Q, Yuanhua H. Reliable imputation of spatial transcriptome with uncertainty estimation and spatial regularization. *bioRxiv*, 2023, 2023.2001.2020.524992
- 79 Li S, Gai K, Dong K, et al. High-density generation of spatial transcriptomics with STAGE. *Nucleic Acids Res*, 2024, 52: 4843–4856
- 80 Rao A, Barkley D, França G S, et al. Exploring tissue architecture using spatial transcriptomics. *Nature*, 2021, 596: 211–220
- 81 Moehlin J, Mollet B, Colombo B M, et al. Inferring biologically relevant molecular tissue substructures by agglomerative clustering of digitized spatial transcriptomes with multilayer. *Cell Syst*, 2021, 12: 694–705.e3
- 82 Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*, 2021, 184: 3573–3587.e29
- 83 Yang Y, Shi X, Liu W, et al. SC-MEB: Spatial clustering with hidden Markov random field using empirical Bayes. *Brief BioInf*, 2022, 23: bbab466
- 84 Patrick E, Canete N P, Iyengar S S, et al. Spatial analysis for highly multiplexed imaging data to identify tissue microenvironments. *Cytometry Pt A*, 2023, 103: 593–599
- 85 Hu J, Li X, Coleman K, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods*, 2021, 18: 1342–1351
- 86 Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun*, 2022, 13: 1739
- 87 Duy P, Xiao T, Jun X, et al. stLearn: Integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *bioRxiv*, 2020, 2020.2005.2031.125658
- 88 Moran P A P. Notes on continuous stochastic phenomena. *Biometrika*, 1950, 37: 17–23
- 89 Xu H, Fu H, Long Y, et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med*, 2024, 16: 12
- 90 Zong Y S, Yu T Y, Wang X S, et al. conST: An interpretable multi-modal contrastive learning framework for spatial transcriptomics. *bioRxiv*, 2022, 2022.2001.2014.476408
- 91 Zong Y S, Yin R, Luo M, et al. Deciphering spatial domains by integrating histopathological image and transcriptomics via contrastive learning. *bioRxiv*, 2022, 2022.2009.2030.510297
- 92 Svensson V, Teichmann S A, Stegle O. SpatialDE: Identification of spatially variable genes. *Nat Methods*, 2018, 15: 343–346
- 93 Edsgård D, Johnsson P, Sandberg R. Identification of spatial expression trends in single-cell gene expression data. *Nat Methods*, 2018, 15: 339–342
- 94 Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat Methods*, 2020, 17: 193–200
- 95 Zhu J, Sun S, Zhou X. SPARK-X: Non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biol*, 2021, 22: 184

- 96 Li Q, Zhang M, Xie Y, et al. Bayesian modeling of spatial molecular profiling data via Gaussian process. *Bioinformatics*, 2021, 37: 4129–4136
- 97 Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 2014, 15: 550
- 98 Robinson M D, McCarthy D J, Smyth G K. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 2010, 26: 139–140
- 99 Abrar M A, Kaykobad M, Rahman M S, et al. NoVaTeST: Identifying genes with location-dependent noise variance in spatial transcriptomics data. *Bioinformatics*, 2023, 39: btad372
- 100 Zhang K, Feng W, Wang P. Identification of spatially variable genes with graph cuts. *Nat Commun*, 2022, 13: 5488
- 101 Hao M, Hua K, Zhang X, et al. SOMDE: A scalable method for identifying spatially variable genes with self-organizing map. *Bioinformatics*, 2021, 37: 4392–4398
- 102 Zhu J, Sabatti C. Integrative spatial single-cell analysis with graph-based feature learning. *bioRxiv*, 2020, 2020.08.12.248971
- 103 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
- 104 Seal S, Bitler B G, Ghosh D, et al. SMASH: Scalable Method for Analyzing Spatial Heterogeneity of genes in spatial transcriptomics data. *PLoS Genet*, 2023, 19: e1010983
- 105 Vandenbon A, Diez D. A universal differential expression prediction tool for single-cell and spatial genomics data. *bioRxiv*, 2022, 2022.11.13.516355
- 106 Syed V. Disruption of germ cell–Sertoli cell interactions leads to spermatogenic defects. *Mol Cell Endocrinol*, 2002, 186: 155–157
- 107 Kamińska K, Szczalik C, Bielecka Z F, et al. The role of the cell–cell interactions in cancer progression. *J Cell Mol Medi*, 2015, 19: 283–296
- 108 Floudas A, Smith C M, Tynan O, et al. Distinct stromal and immune cell interactions shape the pathogenesis of rheumatoid and psoriatic arthritis. *Ann Rheum Dis*, 2022, 81: 1224–1242
- 109 Almet A A, Cang Z, Jin S, et al. The landscape of cell–cell communication through single-cell transcriptomics. *Curr Opin Syst Biol*, 2021, 26: 12–23
- 110 Kumar M P, Du J, Lagoudas G, et al. Analysis of single-cell RNA-Seq identifies cell-cell communication associated with tumor characteristics. *Cell Rep*, 2018, 25: 1458–1468.e4
- 111 Wang Y, Wang R, Zhang S, et al. iTALK: An R package to characterize and illustrate intercellular communication. *bioRxiv*, 2019, 507871
- 112 Cillo A R, Kürten C H L, Tabib T, et al. Immune landscape of viral- and carcinogen-driven head and neck cancer. *Immunity*, 2020, 52: 183–199.e9
- 113 Jin S, Guerrero-Juarez C F, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*, 2021, 12: 1088
- 114 Zhang Y, Liu T, Wang J, et al. Cellinker: A platform of ligand–receptor interactions for intercellular communication analysis. *Bioinformatics*, 2021, 37: 2025–2032
- 115 Choi H, Sheng J, Gao D, et al. Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep*, 2015, 10: 1187–1201
- 116 Browaeys R, Saelens W, Saeys Y. NicheNet: Modeling intercellular communication by linking ligands to target genes. *Nat Methods*, 2020, 17: 159–162
- 117 Cheng J, Zhang J, Wu Z, et al. Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. *Brief BioInf*, 2021, 22: 988–1005
- 118 Zhang Y, Liu T, Hu X, et al. CellCall: Integrating paired ligand–receptor and transcription factor activities for cell–cell communication. *Nucleic Acids Res*, 2021, 49: 8520–8534
- 119 Hu Y, Peng T, Gao L, et al. CytoTalk: De novo construction of signal transduction networks using single-cell transcriptomic data. *Sci Adv*, 2021, 7: eabf1356
- 120 Sheinin R, Dulberg S, Kaminitz A, et al. Identification of factors mediating the signaling convergence of multiple receptors following cell-cell interaction. *bioRxiv*, 2023
- 121 Peng L, Wang F, Wang Z, et al. Cell–cell communication inference and analysis in the tumour microenvironments from single-cell transcriptomics: Data resources and computational strategies. *Brief BioInf*, 2022, 23: bbac234
- 122 Shao X, Lu X, Liao J, et al. New avenues for systematically inferring cell-cell communication: Through single-cell transcriptomics data. *Protein Cell*, 2020, 11: 866–880
- 123 Kleino I, Frolovaitė P, Suomi T, et al. Computational solutions for spatial transcriptomics. *Comput Struct Biotechnol J*, 2022, 20: 4870–4884
- 124 Longo S K, Guo M G, Ji A L, et al. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat Rev Genet*, 2021, 22: 627–644
- 125 Garcia-Alonso L, Handfield L F, Roberts K, et al. Mapping the temporal and spatial dynamics of the human endometrium *in vivo* and *in vitro*. *Nat Genet*, 2021, 53: 1698–1711
- 126 Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat Commun*, 2020, 11: 2084

- 127 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
- 128 Li Z, Wang T, Liu P, et al. SpatialDM for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns. *Nat Commun*, 2023, 14: 3995
- 129 Jerby-Arnon L, Regev A. DIALOGUE maps multicellular programs in tissue from single-cell or spatial transcriptomics data. *Nat Biotechnol*, 2022, 40: 1467–1477
- 130 Armingol E, Ghaddar A, Joshi C J, et al. Inferring a spatial code of cell-cell interactions across a whole animal body. *PLoS Comput Biol*, 2022, 18: e1010715
- 131 Miller B F, Bambah-Mukku D, Dulac C, et al. Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome Res*, 2021, 31: 1843–1855
- 132 Canete N P, Iyengar S S, Ormerod J T, et al. spicyR: Spatial analysis of *in situ* cytometry data in R. *Bioinformatics*, 2022, 38: 3099–3105
- 133 Yuan Y, Bar-Joseph Z. GCNG: Graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol*, 2020, 21: 300
- 134 Li D, Ding J, Bar-Joseph Z, et al. Identifying signaling genes in spatial single-cell expression data. *Bioinformatics*, 2021, 37: 968–975
- 135 Fischer D S, Schaar A C, Theis F J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat Biotechnol*, 2023, 41: 332–336
- 136 Tanevski J, Flores R O R, Gabor A, et al. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biol*, 2022, 23: 97
- 137 Arnol D, Schapiro D, Bodenmiller B, et al. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell Rep*, 2019, 29: 202–211.e6
- 138 Li D, Chen J, Ma Q, et al. Counterfactual hypothesis testing of tumor microenvironment scenarios through semantic image synthesis. *bioRxiv*, 2021, 2020.10.27.358101
- 139 Yáñez-Mó M, Siljander P R, Andreu Z, et al. Biological properties of extracellular vesicles and their physiological functions. *J Extracell Vesicle*, 2015, 4: 27066
- 140 Dimitrov D, Türei D, Garrido-Rodriguez M, et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat Commun*, 2022, 13: 3224
- 141 Wang S, Zheng H, Choi J S, et al. A systematic evaluation of the computational tools for ligand-receptor-based cell–cell interaction inference. *Brief Funct Genomics*, 2022, 21: 339–356
- 142 Li R, Yang X. De novo reconstruction of cell interaction landscapes from single-cell spatial transcriptome data with DeepLinc. *Genome Biol*, 2022, 23: 124
- 143 Sang-aram C, Browaeys R, Seurinck R, et al. Spotless: A reproducible pipeline for benchmarking cell type deconvolution in spatial transcriptomics. *eLife* 2024, 12: RP88431
- 144 Schapiro D, Jackson H W, Raghuraman S, et al. histoCAT: Analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat Methods*, 2017, 14: 873–876
- 145 Wei R, He S, Bai S, et al. Spatial charting of single-cell transcriptomes in tissues. *Nat Biotechnol*, 2022, 40: 1190–1199
- 146 Chen J, Luo T, Jiang M, et al. Cell composition inference and identification of layer-specific spatial transcriptional profiles with POLARIS. *Sci Adv*, 2023, 9: eadd9818
- 147 Liu W, Liao X, Luo Z, et al. Probabilistic embedding, clustering, and alignment for integrating spatial transcriptomics data with PRECAST. *Nat Commun*, 2023, 14: 296
- 148 Zhu J, Shang L, Zhou X. SRTsim: Spatial pattern preserving simulations for spatially resolved transcriptomics. *Genome Biol*, 2023, 24: 39
- 149 Yang J, Qu Y, Wei N, et al. Spider: A flexible and unified framework for simulating spatial transcriptomics data. *bioRxiv*, 2023. doi.org/10.1101/2023.05.21.541605
- 150 Song D, Wang Q, Yan G, et al. scDesign3 generates realistic *in silico* data for multimodal single-cell and spatial omics. *Nat Biotechnol*, 2024, 42: 247–252
- 151 Feng Y, Yang T, Zhu J, et al. Spatial analysis with SPIAT and spaSim to characterize and simulate tissue microenvironments. *Nat Commun*, 2023, 14: 2697
- 152 Eng C H L, Lawson M, Zhu Q, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, 2019, 568: 235–239
- 153 Long Y, Ang K S, Li M, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat Commun*, 2023, 14: 1155
- 154 Xu C, Jin X, Wei S, et al. DeepST: Identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Res*, 2022, 50: e131
- 155 Wang G, Zhao J, Yan Y, et al. Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. *Nat Mach Intell*, 2023, 5: 1200–1213
- 156 Asish Kumar S, Vrushali P, Jyoti S, et al. SpatialPrompt: Spatially aware scalable and accurate tool for spot deconvolution and clustering in spatial

- transcriptomics. *bioRxiv*, 2023, 2023.2009.2007.556641
- 157 Xu H, Wang S, Fang M, et al. SPACEL: Deep learning-based characterization of spatial transcriptome architectures. *Nat Commun*, 2023, 14: 7603
- 158 Zhou X, Dong K, Zhang S. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nat Comput Sci*, 2023, 3: 894–906
- 159 Guo T, Yuan Z, Pan Y, et al. SPIRAL: Integrating and aligning spatially resolved transcriptomics data across different experiments, conditions, and technologies. *Genome Biol*, 2023, 24: 241
- 160 Xia C R, Cao Z J, Tu X M, et al. Spatial-linked alignment tool (SLAT) for aligning heterogenous slices. *Nat Commun*, 2023, 14: 7236
- 161 Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. *Adv Neural Inform Proc Syst*, 2017, 30: 1025–1035
- 162 He B, Bergenstråhle L, Stenbeck L, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng*, 2020, 4: 827–834
- 163 Tan X, Su A, Tran M, et al. SpaCell: Integrating tissue morphology and spatial gene expression to predict disease cells. *Bioinformatics*, 2020, 36: 2293–2294
- 164 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
- 165 Wu Y, Zhou J-Y, Yao B, et al. STASCAN deciphers fine-resolution cell-distribution maps in spatial transcriptomics by deep learning. *bioRxiv*, 2024. Doi: 10.1101/2023.09.02.556029
- 166 Cao Z J, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat Biotechnol*, 2022, 40: 1458–1466
- 167 Long Y H, Siong Ang K, Liao S, et al. Integrated analysis of spatial multi-omics with SpatialGlue. *bioRxiv*, 2023, 2023.2004.2026.538404
- 168 Wang Z Q, Czerniak B, Wei P. Spatial IMIX: A mixture model approach to spatially correlated multi-omics data integration. *bioRxiv*, 2023, 2023.2007.2015.549148
- 169 Zheng Y, Chen Y, Ding X, et al. Aquila: A spatial omics database and analysis platform. *Nucleic Acids Res*, 2023, 51: D827–D834
- 170 Wang G, Wu S, Xiong Z, et al. CROST: A comprehensive repository of spatial transcriptomics. *Nucleic Acids Res*, 2024, 52: D882–D890
- 171 Yuan Z, Pan W, Zhao X, et al. SODB facilitates comprehensive exploration of spatial omics data. *Nat Methods*, 2023, 20: 387–399
- 172 Fan Z, Luo Y, Lu H, et al. SPASER: Spatial transcriptomics annotation at single-cell resolution. *Nucleic Acids Res*, 2023, 51: D1138–D1149
- 173 Marconato L, Palla G, Yamauchi K A, et al. SpatialData: An open and universal data framework for spatial omics. *Nat Methods*, 2024
- 174 Li X, Xiao C, Qi J, et al. STellaris: A web server for accurate spatial mapping of single cells based on spatial transcriptomics data. *Nucleic Acids Res*, 2023, 51: W560–W568
- 175 Xu Z, Wang W, Yang T, et al. STOmicsDB: A comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Res*, 2024, 52: D1053–D1061
- 176 Ji A L, Rubin A J, Thrane K, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, 2020, 182: 1661–1662
- 177 Maynard K R, Collado-Torres L, Weber L M, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci*, 2021, 24: 425–436
- 178 Maniatis S, Åijö T, Vickovic S, et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science*, 2019, 364: 89–93
- 179 Vickovic S, Lötstedt B, Klughammer J, et al. SM-Omics is an automated platform for high-throughput spatial multi-omics. *Nat Commun*, 2022, 13: 795
- 180 Deng Y, Bai Z, Fan R. Microtechnologies for single-cell and spatial multi-omics. *Nat Rev Bioeng*, 2023, 1: 769–784
- 181 Vandereyken K, Sifrim A, Thienpont B, et al. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet*, 2023, 24: 494–515
- 182 Maseda F, Cang Z, Nie Q. DEEPsc: A deep learning-based map connecting single-cell transcriptomics and spatial imaging data. *Front Genet*, 2021, 12: 636743
- 183 Adossa N, Khan S, Rytönen K T, et al. Computational strategies for single-cell multi-omics integration. *Comput Struct Biotechnol J*, 2021, 19: 2588–2596
- 184 Miao Z, Humphreys B D, McMahon A P, et al. Multi-omics integration in the age of million single-cell data. *Nat Rev Nephrol*, 2021, 17: 710–724
- 185 Cui H, Wang C, Maan H, et al. scGPT: Towards Building a Foundation Model for Single-Cell Multi-omics Using Generative AI. *bioRxiv*, 2023
- 186 Qin L, Ding X, Li Y, et al. Co-mutation modules capture the evolution and transmission patterns of SARS-CoV-2. *Brief BioInf*, 2021, 22: bbaa222
- 187 Wang M, Gan J, Han C, et al. Imputation Methods for scRNA Sequencing Data. *Appl Sci*, 2022, 12: 10684
- 188 Pain C D, Egan G F, Chen Z. Deep learning-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement. *Eur J Nucl Med Mol Imag*, 2022, 49: 3098–3118

Summary for “空间转录组的解读与探究: 数据分析方法学研究现状与展望”

Decoding spatial transcriptomics: Current trends and future prospects in data analysis methodologies

Runze Li^{†*}, Xu Chen[†] & Xuerui Yang^{*}

MOE Key Laboratory of Bioinformatics, Center for Synthetic & Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China

[†] Equally contributed to this work

* Corresponding authors, E-mail: lirz2023@mail.tsinghua.edu.cn; yangxuerui@tsinghua.edu.cn

In recent years, spatial transcriptomics has emerged as a rapidly advancing technology, fundamentally transforming research paradigms across various fields of biology and medicine. This innovative technology retains the spatial localization information of cells within complex tissues and performs transcriptome analysis at multicellular, single-cell, or subcellular resolution. By coupling spatial localization information with molecular profiles, spatial transcriptomics generates new types of multimodal high-throughput data. These data sets serve as insightful resources for exploring molecular and cellular mechanisms within physiological contexts. For instance, spatial transcriptomics can elucidate the intricate architecture of the tumor microenvironment, revealing how cancer cells interact with surrounding stromal and immune cells. Similarly, it can map the spatial organization of neural circuits in the brain, offering deeper insights into neural development, function, and disease. This capacity to link cellular function to precise tissue architecture marks a significant advancement over traditional transcriptomics, which often lacks spatial context.

Despite its transformative potential, spatial transcriptomic data exhibit several highly non-ideal characteristics, including low signal-to-noise ratio, high sparsity, and uneven coverages. These inherent challenges pose significant obstacles to in-depth data analysis and information mining. The low signal-to-noise ratio often results from technical limitations in capturing and amplifying the RNA transcripts, leading to significant background noise that can obscure true biological signals. High sparsity is a consequence of the limited sensitivity of current technologies, where many genes may not be detected in all cells, resulting in numerous zero counts in the data matrix. Uneven coverage refers to the inconsistent detection of transcripts across different regions of the tissue, which can bias the analysis and interpretation of spatial patterns.

These challenges necessitate the development of sophisticated computational methods to preprocess and analyze the data effectively. For instance, denoising algorithms are essential to enhance the signal quality, enabling more accurate downstream analyses. Imputation techniques aim to address the issue of sparsity by predicting missing values, thereby providing a more complete picture of the transcriptomic landscape. Furthermore, normalization strategies are required to correct for uneven coverage, ensuring that comparisons across different regions and samples are valid. Many of these challenges remain unresolved, complicating efforts to fully leverage the technology's capabilities. Furthermore, the rapidly evolving nature of spatial transcriptomics necessitates continuous adaptation of existing methodologies and the development of more innovative analytical tools to keep pace with technological advancements. As new platforms and techniques are introduced, there is a constant need for updating analytical pipelines and validation frameworks to ensure robust and reproducible results. The integration of spatial transcriptomics with other omics data, such as genomics, proteomics, and metabolomics, also requires the development of novel multi-omics analysis methods that can handle the increased complexity and data dimensionality.

spatial transcriptomics, bioinformatics, artificial intelligence, machine learning, data mining

doi: [10.1360/TB-2024-0332](https://doi.org/10.1360/TB-2024-0332)