基于变分自编码器的空间转录组 细胞聚类研究

刘 腾,李鑫,印明柱*

(重庆大学附属三峡医院,重庆 万州 404000)

摘 要:空间转录组测序技术能够在生成基因表达谱的同时,保留细胞在组织内部的位置信息。如何充分利用基因表达谱和 空间位置信息来识别空间区域,完成细胞亚群聚类是空间转录组学数据分析的基础和关键。本文提出基于变分自编码器和 图神经网络结合的空间转录组细胞亚群聚类方法。构建双层编码器结构,每一层包含简化图卷积(Simple graph convolution, SGC),用以生成低维表征。解码器用以重构特征矩阵,通过最小化损失函数来提高低维表征质量。对低维表征进行下游聚 类,生成不同的细胞亚群。本文提出的聚类方法与多个基准方法在常用的空间转录数据集上进行比较,在聚类准确性和适应 性方面都有优势,证明了该方法的有效性。

关键词:空间转录组学;变分自编码器;图神经网络;细胞聚类 中图分类号:Q2 文献标志码:A 文章编号:1672-5565(2024)04-270-07

Variational autoencoder enabled cell clustering method for spatial transcriptomics

LIU Teng, LI Xin, YIN Mingzhu*

(Chongqing University Three Gorges Hospital, Wanzhou 404000, Chongqing, China)

Abstract: Spatial resolved transcriptomics technology can simultaneously generate gene expression profiles while preserving the positional information of cells within the tissue. How to fully utilize gene expression profiles and spatial positional information to identify spatial regions and complete cell subpopulation clustering is the basis and key for spatial transcriptomics data analysis. In this paper, a spatial transcriptomics cell clustering method based on the combination of variational autoencoder and graph neural network is presented. A two-layer encoder structure is constructed, with each layer containing Simple graph convolution (SGC) to generate low-dimensional representations. The decoder is used to reconstruct the feature matrix and improve the quality of low-dimensional representations to generate different cell subpopulations. The proposed clustering method is compared with several benchmark methods on multiple datasets and has advantages in clustering accuracy and adaptability, demonstrating the effectiveness of the proposed method.

Keywords: Spatial transcriptomics; Variational autoencoder; Graph neural network; Cell clustering

单细胞测序技术能够对单个细胞的基因组、转 录组、蛋白组和代谢组进行测量^[1],进而揭示细胞 类型、罕见亚型、功能状态以及发育轨迹等信息,为 疾病诊断、治疗和预防提供更有效的方法和手段。

引用格式:刘腾,李鑫,印明柱.基于变分自编码器的空间转录组细胞聚类研究[J].生物信息学,2024,22(4):270-276.

收稿日期:2023-05-09;修回日期:2023-07-15;网络首发日期:2023-10-09.

网络首发地址:https://link.cnki.net/urlid/23.1513.q.20231008.1411.002

基金项目:科技部重点研发基金项目(No. 2022YFC3601800);重庆市教育委员会科学技术研究计划重大项目(No. KJZD-K202300105);重庆大学附属三峡医院基础医学重点项目(No. 2023YJKYXML-001).

^{*} 通信作者:印明柱,男,教授,研究方向:生物信息学. E-mail: yinmingzhu2008@126.com.

LIU Teng, LI Xin, YIN Mingzhu. Variational autoencoder enabled cell clustering method for spatial transcriptomics [J]. Chinese Journal of Bioinformatics, 2024, 22(4):270-276.

单细胞测序技术还能探索肿瘤内部的异质性和细胞 微环境,揭示免疫细胞在不同病理状态下的分子机 制和亚群特征,开发更有效的免疫治疗策略^[2]。

空间转录组学技术可以在揭示组织或细胞基因 表达谱的同时,保留细胞的空间位置信息^[3]。通过 对基因表达谱进行高分辨率检测并保留测序细胞的 空间坐标,有助于探索基因表达的时空动态变化,从 而更好地理解细胞功能和组织结构、研究细胞和组 织的异质性、探明疾病的发生机理和进展过程^[4]。 目前,国内外已有多种空间转录技术,如 SlideseqV2^[5], LCM-seq^[6], Stereo-seq^[7] 以及 10x Visium^[8]。

细胞亚群聚类和空间域识别是空间转录组学数 据分析的关键。空间域识别和细胞亚群聚类指将空 间位置信息与基因表达谱数据相结合,鉴定不同生 理状态下具有类似表达模式的细胞,推断细胞类型 及其相互作用方式^[9]。基于识别的空间区域和细 胞亚群,能够开展刻画基因表达模式和基因增补,空 间位点去卷积,细胞空间位置重构,细胞通讯和基因 交互等研究。

目前,国内外学者开展了一系列空间转录组细胞亚群聚类和空间域识别的研究。如纽约大学基因 组学与系统生物学中心发布基于 R 语言的 Seurat 工具箱^[10],将单细胞测序和空间转录组学数据分析 流程化,包含质控、细胞筛选、细胞类型鉴定、特征基 因选择、差异表达分析、数据可视化等功能。德国环 境健康研究中心计算生物学研究所发布基于 Python 的 Scanpy 和 Squidpy 工具箱^[11-12],构建了在 Python 中单细胞测序和空间转录组学数据分析生态 环境,包含了数据预处理、可视化、聚类、伪时间分 析,轨迹推断和差异表达分析等手段。在 Seurat 和 Scanpy 中,主要细胞亚群聚类和空间域识别的方法 是 K-means^[13]和 Leiden^[14],它们都是传统的聚类 算法。mclust^[15]是利用 R 语言编写的基于高斯混 合模型的聚类方法,在处理具有不同分布的数据时, 具有很好的效果。隐马尔可夫随机场^[16](Hidden markov random field,HMRF)模型能够有效地描述细 胞空间位置上下文信息,通过构建能量函数来实现 细胞类型识别。全贝叶斯统计模型^[17](Fully bayesian statistical model)定义基于最大似然估计的 聚类优化目标函数,根据细胞亚群划分间的加权似 然进行聚类。然而,上述聚类方法只考虑了细胞基 因表达谱信息,并没有兼顾空间位置信息,导致空间 域识别和细胞亚群聚类的结果精度并不高。

近两年,图深度学习被认为是分析空间转录组 学数据的有效方法^[18]。图神经网络能充分利用空 间转录组数据中的细胞基因表达谱和空间位置信 息,因此它是提升空间转录组细胞亚群聚类精度的 强力手段。本文提出基于变分自编码器和图神经 网络的细胞亚群聚类方法,如图1所示。通过基 因表达谱筛选高度可变基因,组成特征矩阵;利用 空间位置信息计算欧式距离,形成邻接矩阵。特 征矩阵和邻接矩阵经变分自编码器处理,生成低 维表征,下游聚类方法对低维表征进行分类,得到 不同细胞亚群。其中,变分自编码器由编码器和 解码器组成,均是双层结构,每一层包含简化图卷 积(Simple graph convolution, SGC),用以生成低维 表征。在常用空间转录数据集上,将提出的分类 方法与多个基准方法进行比较,证明其具有更好 的聚类准确性。



图 1 基于变分自编码器的空间转录组细胞亚群聚类架构

Fig.1 Cell clustering architecture for spatial transcriptomics based on variational autoencoder

1 材料与方法

1.1 变分自编码器(VGAE)

在细胞亚群聚类任务中,变分自编码器 (VGAE)的输入是由基因表达谱生成的特征矩阵 *X* 和由空间位置信息计算得到的邻接矩阵 *A*。变分 自编码器的关键点是应用双层图神经网络结构来生 成低维表征。第一层图神经网络用以计算一个低维

特征矩阵 $\widetilde{X}^{[19]}$:

$$\tilde{X} = GNN(X, A) = \operatorname{Re}LU(A X W_0)$$

$$\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$
(1)

其中, \hat{A} 是对称归一化邻接矩阵, ReLu(.) = max(0,.) 是整流线性激活函数, W_0 是这一层图神经网 的权重参数。

在第二层图神经网络中,使用权重参数 W₁计 算特征矩阵的均值和方差向量:

$$\mu = GNN_{\mu}(X, A) = \mathring{A} \mathring{X} W_{1}$$

$$\log \sigma^{2} = GNN(X, A) = \mathring{A} \mathring{X} W_{2}$$
(2)

其中,均值和方差共享同一类权重系数。在此基础 上,利用重参数化方法^[19]来计算得到低维表征:

 $Z = \mu + \sigma e \varepsilon$ (3) $\epsilon \in Norm(0,1)$ 表示标准正态分布。在这里,解码 器用矩阵内积来实现,因而邻接矩阵重构为

$$p(A \mid Z) = \sigma(Z \cdot Z^{T})$$
(4)

最后,损失函数包含两种类型的误差。第一类是重 构误差,用于衡量输入和输出邻接矩阵之间的相似 性。第二类误差使初始标签 q 和预测标签 p 尽可能 接近。损失函数的数学表达式为

$$L = \mathbf{E}_{q(Z|X,A)} \left[\log p(A \mid Z) \right] - KL(q(Z \mid X, A) \| p(Z))$$
(5)

其中, *KL*(.) 表示两个概率分布之间的 Kullback-Leibler 散度。最后,对生成的低维表征进行下游聚 类,获得细胞亚群聚类标签。聚类标签与金标准进 行比较,验证聚类方法的准确性。

1.2 图神经网络

本文采用简化图卷积(Simple graph convolution, SGC)嵌入到变分自编码器中,用以计算低维表征。 简化图卷积在原有的图卷积神经网络(GCN)^[20]上 进行修改,以修正模型复杂性和冗余计算。通过减 少连续层之间的折叠权重矩阵和非线性来简化计 算,提升效率。这种简化的线性模型在理论和实验 中被证明具有更好的性能,其中的卷积核被修改为 线性模式:

$$\hat{Y}_{SGC} = softmax(S...SSX\Theta^{(1)}\Theta^{(2)}...\Theta^{(K)})$$

$$softmax(S^{\kappa}X\Theta)$$
 (6)

其中S是归一化邻接矩阵,X是特征矩阵, Θ 是权重矩阵,softmax表示归一化指数函数。

1.3 空间转录组验证数据集

本文采用最常用的空间转录组学数据集来验证 细胞亚群聚类算法的准确性。这个数据集是关于人 类背外侧前额叶皮层(Human DLPFC),原始数据可 在 10x Visium 数据存储库中获取,包括基因表达谱、 空间位置信息和组织学图像。该数据集相应的细胞 亚群注释金标准可以在 SpatialLIBD 项目^[21]中 找到。

该数据集包含 12 个组织切片,每个切片的捕捉 位点从3 460到4 789不等,测序的基因数量为33 538 个。样本名称标记为151 507到151 676。在人工注 释中,每个样本包含五或七个细胞亚群,包含两类细 胞,即 DLPFC 层和白质。每个细胞亚群具有清晰的 边界,因此这个数据集非常适合用于聚类准确度评 估。将提出的聚类方法与多个基准方法在这个数据 集上进行比较,以评估它们的聚类准确性。此外,每 个样本包含的细胞亚群均按时间顺序排列的,因此 该数据集也适合空间轨迹推断评估。

1.4 评价指标

在空间转录组数据集拥有人工注释的情况下, 可以通过调整兰德指数(ARI)来评估细胞亚群聚类 的准确性。本文使用的人类背外侧前额叶皮层拥有 金标准,因此通过融合预测标签和金标准来计算调 整兰德指数。设 $P = \{P_1, P_2, \cdots P_c\}$ 和 $G = \{G_1, G_2, \cdots, G_c\}$ 分别为预测标签和人工注释标签,则调整兰 德指数可以由下面的公式计算得到:

ARI =

$$\frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_{i} \binom{n_{i}}{2} \sum_{j} \binom{n_{j}}{2}\right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_{i} \binom{n_{i}}{2} + \sum_{j} \binom{n_{j}}{2}\right] - \left[\sum_{i} \binom{n_{i}}{2} \sum_{j} \binom{n_{j}}{2}\right] / \binom{n}{2}}$$
(7)

其中, c 代表细胞亚群的个数, n_i, 和 n_j表示属于第 i 类预测标签和第 j 类人工注释标签的个数, n_{ij} 则是 同时属于第 i 类预测标签和第 j 类人工注释标签的 细胞个数。调整兰德指数的取值范围是[0,1],在 本文中由 scikit-learn 工具包中的 adjusted_rand_ score 函数来计算,其值越高,则细胞亚群聚类的精 度越高。

1.5 神经网络超参数的选取

基于变分自编码器的聚类方法是在 Python 中

利用 PyTorch_pyG^[22]实现。首先,在 Scanpy 工具 箱^[23]中对基因表达谱进行标准化和对数转换,遴选 3 000 个高度可变基因来得到特征矩阵;然后,利用 scikit-learn 工具包^[24]中的最近邻搜索技术来计算 邻接矩阵,每个细胞的邻居通过 K 最近邻法来获 取,其中 K 值取 6。接着,将特征矩阵和邻接矩阵输 入到构建的变分自编码器中生成低维表征。相关的 超参数定义如下:输入维度为 3 000(高变基因数), 隐藏维度为 128,输出维度为 30。学习率为 1×10^{-6} , 迭代次数为 5 000,权重衰减因子为 1×10^{-4} ,梯度剪 裁为 5,随机种子为零。隐藏层的数量为 4,每个层 外的激活函数为指数线性单元(ELU)函数。本文采 用的硬件平台规格为 Ubuntu 20.04.5 LTS 系统搭配 i9-12900F CPU,64G 内存和 GeForce RTXTM 3090Ti GPU。

计算得到低维表征后,采用传统的 Kmeans 聚 类方法来聚类细胞亚群,并在 scikit-learn 软件包中 实现。Kmeans 方法中,细胞亚群的个数设置为与金 标准的个数相同,得到预测细胞亚群标签。最后,根 据预测细胞标签和人工注释标签,计算调整兰德指 数,比较调整兰德指数的大小来说明各种方法的准 确性。本文用到的基准方法包括 Scanpy^[23], SpaGCN^[25], SEDR^[26], STAGATE^[27], DeepST^[28], BayesSpace^[29]等,它们均是先进的基于图神经网络 的空间转录组细胞亚群聚类和空间域识别算法。

2 结果与讨论

2.1 聚类准确度比较

本文提出的细胞亚群聚类和空间域识别方法命 名为 VGAE_SGC。将 VGAE_SGC 与六种空间聚类 基准方法 (Scanpy, SpaGCN, SEDR, STAGATE, DeepST, BayesSpace)进行比较,来说明其准确性和 有效性。将上述七种方法在人类背外侧前额叶皮层 (Human DLPFC)空间转录数据集上进行比较,计算 每种方法在 12 个样本上的调整兰德指数(ARI),生 成 ARI 平均值。

图 2 展示了每种方法在 Human DLPFC 数据上 的调整 兰德指数 平均值,可以看到本文提出的 VGAE_SGC 方法拥有最大的 ARI (0.539)。在七种 方法中,有三种方法的 ARI 平均值超过了 0.5,分别 是 VGAE_SGC, DeepST 和 STAGATE。同时, Scanpy 中采用传统的 Leiden 算法进行细胞亚群聚类,并没 有利用空间转录数据中的空间位置信息,因此它对 应的平均 ARI 值最低,为 0.311。SpaGCN, SEDR 和 BayesSpace 的精度处于中间位置, ARI 值均在 0.45 左右。图 2 说明本文提出的方法在聚类精度方面能 够优于存在的细胞亚群聚类方法。





2.2 空间域识别结果比较

为了进一步说明 VGAE_SGC 在空间域识别方

面的优势,以 Human DLPFC 数据集中的单个样本 为例,来展示七种方法的空间域识别结果。本文选

取样本 151 675 来展示结果,图 3 是每种方法在该 样本上的细胞亚群聚类结果。该样本的金标准 (Ground truth)中包含 7 个空间域,每个空间域之间 的界限明确,而且空间域之间按一定的时间先后顺 序排列。

在该样本上,本文提出的 VGAE_SGC 仍具有最高的 ARI 值,为 0.550。剩余六种基准方法的 ARI 值均没有超过 0.50,以此说明 VGAE_SGC 在这个样

本上的聚类优势。从空间域识别的结果来看, VGAE_SGC 和 DeepST 结果更加接近于金标准,任 何两个空间域之间都有明确的分界线。在 SpaGCN 和 Scanpy 中,由于两者的 ARI 较低,对应它们的空 间域识别结果也比较差,细胞的分布杂乱无章。图 3 说明了本文提出的 VGAE_SGC 能够更好地利用 空间位置信息来定位细胞,从而识别更加准确的空 间域。



图 3 七种方法的空间转录组数据空间域识别结果比较 Fig.3 Spatial domain identification results of seven compared methods for spatial transcriptomics

2.3 低维表征质量比较

在本文比较的七种方法中,四种方法的低维表 征能够获取,从而得到降维聚类图。本节比较四种 方法(VGAE_SGC, DeepST, STAGATE 和 SEDR)的 低维表征的质量。具体的降维图展示在图 4 中。首 先, VGAE_SGC 对应的调整兰德指数最大, 为0.550。 DeepST 比较接近 VGAE SGC, ARI 值为 0.50。它 们对应的 Umap 降维图也比较接近,每个细胞亚群 都独立分开,而且按一定先后顺序排列。而在 STAGATE 和 SEDR 的降维图中,结果就比较差。 比如 STAGATE 中的 Layer_1 和 Layer_2, SEDR 中 的细胞亚群错乱的分布,并没有按一定的先后顺 序来排列,说明它并没有充分结合基因表达谱和 空间位置信息来完成聚类。综上,与已有的空间 转录组细胞亚群聚类基准方法相比,上述三部分 结果证明了本文提出的 VGAE SGC 方法的准确性 和有效性。

3 结 论

利用图深度学习来结合空间转录组中基因表达 谱和空间位置信息来进行细胞亚群聚类能够有效地 提升准确性。本文构建名为 VGAE_SGC 的空间转 录组聚类架构,在变分自编码器中嵌入简化图卷积 来处理基因表达谱和空间位置信息。在人类背外侧 前额叶皮层(Human DLPFC)空间转录数据集上与 多个基准方法进行比较,说明了 VGAE_SGC 方法在 细胞亚群聚类精度、空间域识别的结果和低维表征 质量三个方面都有优势,证明了本文提出的基于变 分自编码器的空间转录组细胞亚群聚类方法的有效 性。本文聚类方法的源代码及其在更多数据集中的 聚 类 结 果 可 见 链 接 https://github.com/ narutoten520/Bioinformatics_VGAE_SGC。



图 4 四种方法中低维表征的 Umap 降维图 Fig.4 The Umap diagram of latent embedding in four methods

参考文献(References)

- [1]NAWY T. Single-cell sequencing [J]. Nature Methods, 2014, 11(1): 18-18. DOI:10.1038/nmeth.2771.
- [2] 卢汀. 生物信息学基因表达差异分析[J]. 生物信息学, 2014,12(2):140-144. DOI: 10.3969/j.issn.1672-5565. 2014-02.20140210.

LU Ting. Bioinformatics analysis for gene differential expression [J]. Chinese Journal of Bioinformatics, 2014, 12(2):140-144. DOI: 10.3969/j.issn.1672-5565.2014-02.20140210.

[3] 许醒, 蔡林峰, 张倩楠, 等. 单细胞与空间转录组分析 研究进展[J]. 分析测试学报, 2022, 41(9):1322-1334. DOI: 10.19969/j.fxcsxb.22052401.

XU Xing, CAI Linfeng, ZHANG Qiannan, et al. Advances in single-cell and spatial transcriptome analysis [J]. Journal of Analytical Testing, 2022, 41(9):1322-1334. DOI: 10. 19969/j.fxcsxb.22052401.

- [4] WILLIAMS C G, LEE H J, ASATSUMA T, et al. An introduction to spatial transcriptomics for biomedical research[J]. Genome Medicine, 2022, 14(1): 1-18. DOI: 10.1186/s13073-022-01075-1.
- [5] RODRIQUES S G, STICKELS R R, GOEVA A, et al. Slide-seq: A scalable technology for measuring genome-wide

expression at high spatial resolution [J]. Science, 2019, 363(6434): 1463-1467. DOI: 10.1126/science.aaw1219.

- [6] NICHTERWITZ S, CHEN G, AGUILA BENITEZ J, et al. Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling[J]. Nature Communications, 2016, 7(1): 12139. DOI: 10.1038/ncomms12139.
- [7]XIA Keke, SUN Haixi, LI Jie, et al. Single-cell Stereo-seq enables cell type-specific spatial transcriptome characterization in Arabidopsis leaves [J]. bioRxiv, 2021; 2021.10. 20. 465066. DOI: 10.1101/2021.10.20.465066.
- [8] RAO N, CLARK S, HABERN O. Bridging genomics and tissue pathology: 10x genomics explores new frontiers with the visium spatial gene expression solution[J]. Genetic Engineering & Biotechnology News, 2020, 40(2): 50-51. DOI: 10.1089/gen.40.02.16.
- [9] 王琳,赵桂华.用于空间分辨转录组学数据分析的统计 方法[J]. Bioprocess, 2023, 13: 57. DOI: 10.12677/BP. 2023.131008.
 WANG Lin, ZHAO Guihua. Statistical methods for spatially re-solved transcriptomic data analysis [J]. Bioprocess,

2023, 13: 57. DOI: 10.12677/BP.2023.131008.

- [10]SATIJA R, FARRELL J A, GENNERT D, et al. Spatial reconstruction of single-cell gene expression data [J]. Nature Biotechnology, 2015, 33(5): 495-502. DOI: 10. 1038/nbt.3192.
- [11] WOLF F A, ANGERER P, THEIS F J. SCANPY: Large-

scale single-cell gene expression data analysis[J]. Genome Biology, 2018, 19: 1-5. DOI: 10.1186/s13059-017-1382-0.

- [12] PALLA G, SPITZER H, KLEIN M, et al. Squidpy: A scalable framework for spatial omics analysis [J]. Nature methods, 2022, 19(2):171-178. DOI: 10.1038/s41592-021-01358-2.
- [13] 宋东奇,宋余庆,刘哲,等. 新型适用于基因表达数据的模型聚类方法[J]. 计算机与应用化学,2015,32(1):71-74. DOI:10.3390/diagnostics10080584.
 SONG Dongqi, SONG Yuqing, LIU Zhe, et al. Novel model clustering approach for gene expression data[J]. Computer and Applied Chemistry, 2015, 32(1):71-74. DOI: 10.3390/diagnostics10080584.
- [14] TRAAG V A, WALTMAN L, VAN ECK N J. From louvain to leiden: Guaranteeing well-connected communities [J]. Scientific Reports, 2019, 9(1): 5233. DOI: 10. 1038/s41598-019-41695-z.
- [15] HAUGHTON D, LEGRAND P, WOOLFORD S. Review of three latent class cluster analysis packages: Latent Gold, poLCA, and MCLUST [J]. The American Statistician, 2009, 63(1): 81-91. DOI: 10.1198/tast.2009.0016.
- [16] CHATZIS S P, TSECHPENAKIS G. The infinite hidden Markov random field model [J]. IEEE Transactions on Neural Networks, 2010, 21(6): 1004-1014. DOI: 10. 1109/TNN.2010.2046910.
- [17] MO Qianxing, SHEN Ronglai, GUO Cui, et al. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data [J]. Biostatistics, 2018, 19(1): 71-86. DOI: 10.1093/biostatistics/kxx017.
- [18] 王国力,孙宇,魏本征. 医学图像图深度学习分割算法
 综述[J]. 计算机工程与应用,2022,58(12):37-50.
 DOI: 10.3778/j.issn.1002-8331.2112-0225.
 WANG Guoli, SUN Yu, WEI Benzheng. Systematic review
 - on graph deep learning in medical image segmentation[J]. Computer Engineering and Applications, 2022, 58(12): 37-50. DOI: 10.3778/j.issn.1002-8331.2112-0225.
- [19] KIPF T N, WELLING M. Variational graph auto-encoders
 [J]. arXiv preprint arXiv:1611.07308, 2016. DOI: 10.
 48550/arXiv.1611.07308.
- [20] ZHANG Si, TONG Hanghang, XU Jiejun, et al. Graph

convolutional networks: A comprehensive review [J]. Computational Social Networks, 2019, 6(1): 1–23. DOI: 10. 1186/s40649-019-0069-y.

- [21] PARDO B, SPANGLER A, WEBER L M, et al. Spatial-LIBD: An R/Bioconductor package to visualize spatiallyresolved transcriptomics data[J]. BMC Genomics, 2022, 23(1): 1-5. DOI: 10.1186/s12864-022-08601-w.
- [22] FEY M, LENSSEN, J E. Fast graph representation learning with PyTorch Geometric [J]. arXiv preprint: 1903. 02428,2019. DOI: 10.48550/arXiv.1903.02428.
- [23] WOLF F A, ANGERER P, THEIS F J. SCANPY: Largescale single-cell gene expression data analysis[J]. Genome Biology, 2018, 19 (1): 1-5. DOI: https://doi.org/10. 1186/s13059-017-1382-0.
- [24] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: Machine learning in Python [J]. The Journal of Machine Learning Research, 2011, 12: 2825-2830. DOI: 10.48550/arXiv.1201.0490.
- [25] HU Jian, LI Xiangjie, COLEMAN K, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network [J]. Nature Methods, 2021, 18(11): 1342-1351. DOI: 10.1038/s41592-021-01255 -8.
- [26] FU Huazhu, XU Hang, CHONG K, et al. Unsupervised spatially embedded deep representation of spatial transcriptomics[J]. Biorxiv, 2021: 2021.06. 15.448542. DOI: 10. 1101/2021.06.15.448542.
- [27] DONG Kangning, ZHANG Shihua. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder[J]. Nature Communications, 2022, 13(1): 1739. DOI: 10.1038/s41467-022-29439-6.
- [28]XU Chang, JIN Xiyun, WEI Songren, et al. DeepST: Identifying spatial domains in spatial transcriptomics by deep learning[J]. Nucleic Acids Research, 2022, 50(22): e131-e131. DOI: 10.1093/nar/gkac901.
- [29] ZHAO E, STONE M R, REN Xing, et al. Spatial transcriptomics at subspot resolution with BayesSpace[J]. Nature Biotechnology, 2021, 39(11): 1375-1384. DOI: 10. 1038/s41587-021-00935-2.