
QB期刊 耶鲁大学赵宏宇教授团队展望空间转录组数据的基础模型

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/39705.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

QB期刊 耶鲁大学赵宏宇教授团队展望空间转录组数据的基础模型。论文标题：A perspective on developing foundation models for analyzing spatial transcriptomic data

期刊：Quantitative Biology

作者：Tianyu Liu, Minsheng Hao, Xinhao Liu, Hongyu Zhao

发表时间：28 Aug 2025

DOI：10.1002/qub2.70010

微信链接：点击此处阅读微信文章

基础模型是一种基于大规模数据预训练的深度学习模型，可应用于多种下游任务，已彻底改变了自然语言处理和多模态人工智能领域。然而，在空间转录组学（ST）领域，目前尚无基础模型展现出生成经验证的新型生物学发现的能力。

近期，美国耶鲁大学赵宏宇教授团队在Quantitative Biology期刊上发表了题为A perspective on developing foundation models for analyzing spatial transcriptomic data的Perspective文章。文章认为，由于ST数据缺乏明确的序列状结构、存在噪声，且其采集成本高于单细胞RNA测序数据，不适合简单地复用现有的单细胞基础模型。因此，如何利用ST数据构建更好的基础模型是非常有前景的研究方向，并需要更多的探索。

A perspective on developing foundation models for analyzing spatial transcriptomic data

Tianyu Liu^{1,2} | Minsheng Hao³ | Xinhao Liu⁴ | Hongyu Zhao^{1,2}

¹Interdepartmental Program of Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, USA

²Department of Biostatistics, Yale University, New Haven, Connecticut, USA

³Research and Early Development, Genentech, South San Francisco, California, USA

⁴Department of Computer Science, Princeton University, Princeton, New Jersey, USA

全文概要

本文将用于ST分析的基础模型分为两类。基于序列的基础模型是通过自监督学习直接在大规模ST测序数据上预训练的，例如NicheCompass、Nicheformer、STFormer和CellPLM。而基于知识的基础模型则利用现有的LLM或预训练于生物文本或病理图像的大型多模态模型（如QuST-LLM和Geneverse），将通用知识迁移到空间分析中。文章还重点介绍了结合这两种范式的混合方法，例如spEMO和scGPT-spatial，如图1所示。

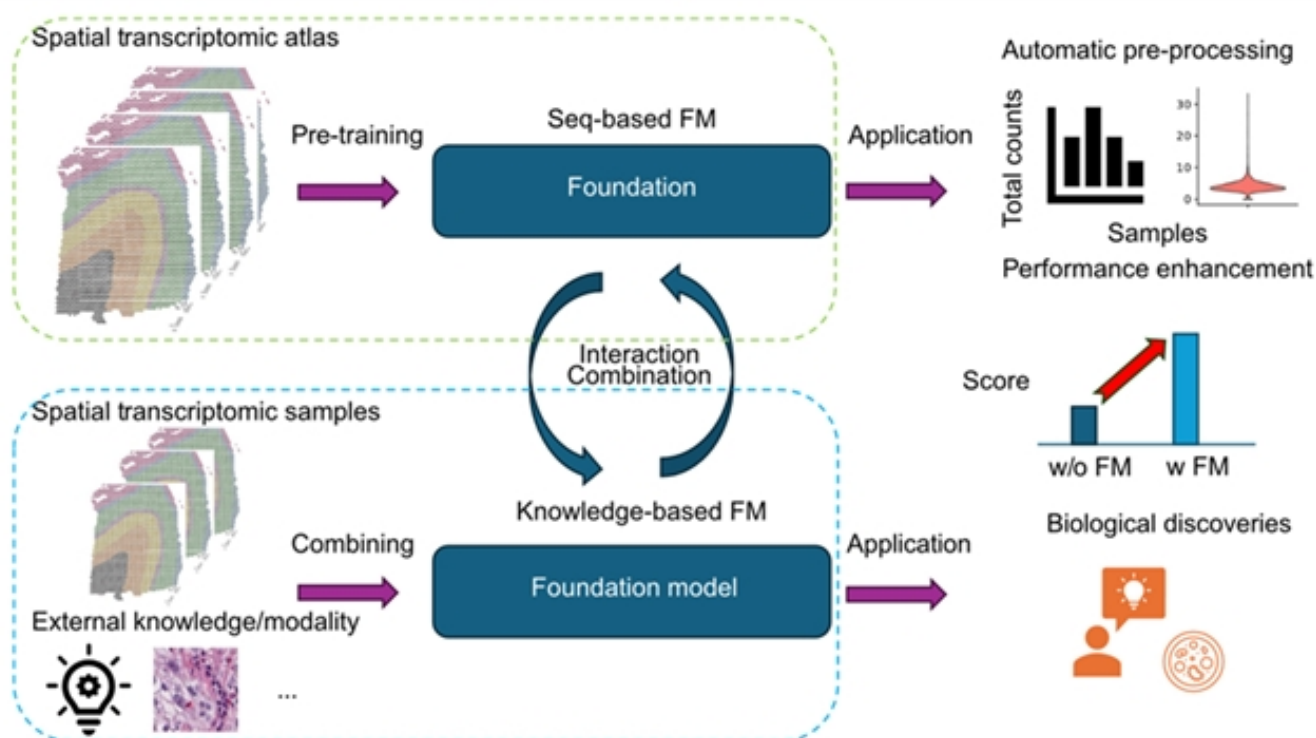


图1. 基于序列和知识的空间转录组学基础模型构建

文章认为，基础模型应致力于解决具有实质性影响的重要问题，而非基础聚类简单任务。具体而言，特征管理应有助于实现预处理流程（包括质量控制、标准化和标注）的自动化与标准化，从而降低主观性，并提高不同研究间的可重复性。此外，特征管理还应提升关键下游任务的性能，例如细胞类型标注、空间生态位聚类、基因表达插补以及空间去卷积。

用于ST分析的基础模型的一大机遇在于加速生物学发现，从而减少对昂贵湿实验的需求。类比化学领域的ChemCrow等工具，一个能够利用基础模型的人工智能代理能够识别新型细胞类型、预测扰动效应，并探索由空间因素诱发的生物学模式。这些领域目前仍处于未开发状态。作者特别强调，必须权衡图论模型的价值与人力资源及模型训练的成本。

展望未来，若要充分释放利用ST开发的基础模型的潜力，必须应对多项挑战。其中包括：收集高质量且多样化的训练数据；设计适用于非序列转录组数据的预训练目标；以及构建超越低层级任务的严谨基准测试框架。此外，还需审慎管理计算成本；作者倡导开源不同规模的模型，并提供在线演示服务，从而使这些工具能够广泛惠及整个科研社区。



QB期刊介绍

Quantitative Biology (QB) 期刊是由清华大学、北京大学、高等教育出版社联合创办的全英文学术期刊。由高等教育出版社和Wiley双平台出版和发行。QB主要刊登生物信息学、计算生物学、系统生物学、理论生物学和合成生物学的最新研究成果和前沿进展，并为生命科学与计算机、数学、物理等交叉研究领域打造一个学术水平高、可读性强、具有全球影响力的交叉学科期刊品牌。

QB期刊目前已被ESCI, PMC, Scopus, DOAJ, CSCD等国内外重要数据库收录。

《前沿》系列英文学术期刊

由教育部主管、高等教育出版社主办的《前沿》(Frontiers) 系列英文学术期刊，于2006年正式创刊，以网络版和印刷版向全球发行。系列期刊包括基础科学、生命科学、工程技术和人文社会

科学四个主题，是我国覆盖学科最广泛的英文学术期刊群，其中12种被SCI收录，其他也被A&HCI、Ei、MEDLINE或相应学科国际权威检索系统收录，具有一定的国际学术影响力。系列期刊采用在线优先出版方式，保证文章以最快速度发表。

中国学术前沿期刊网

<http://journal.hep.com.cn>



来源：Quantitative Biology

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发