CellOntologyMapper：细胞类型注释的共识映射

曾泽华[1,2,3#]，王学海[4,5#]，杜宏武[1,2*]，幸岑璨[1,2*]

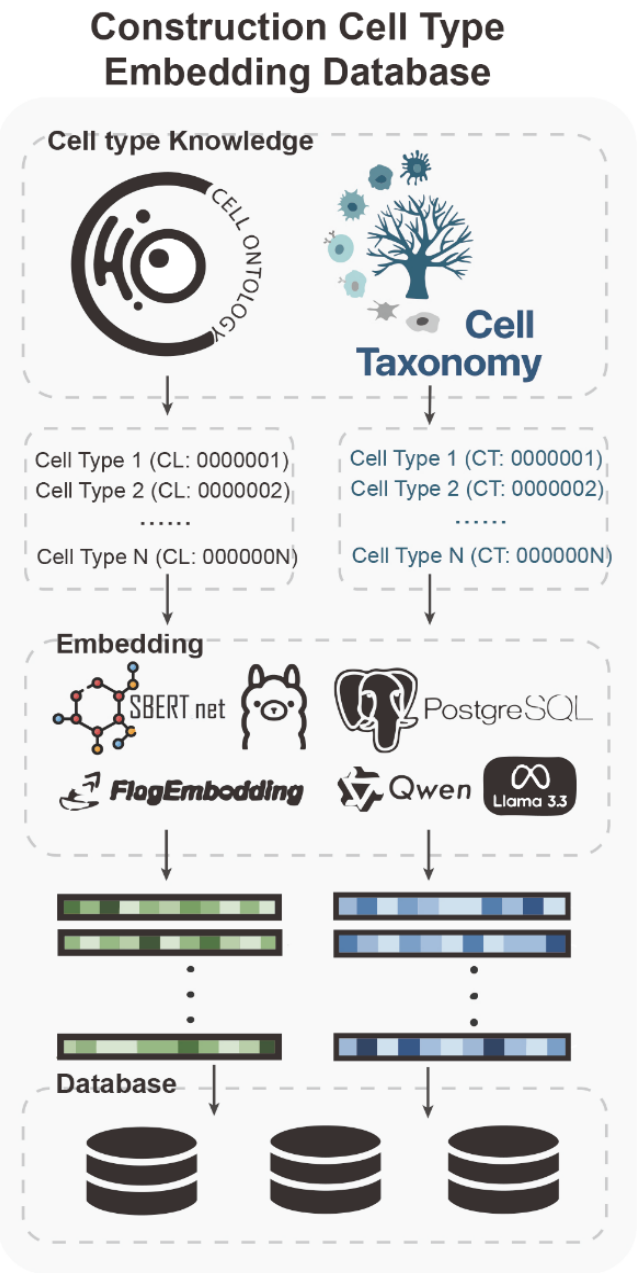
[1]北京科技大学大兴研究院 [2]北京科技大学化生学院

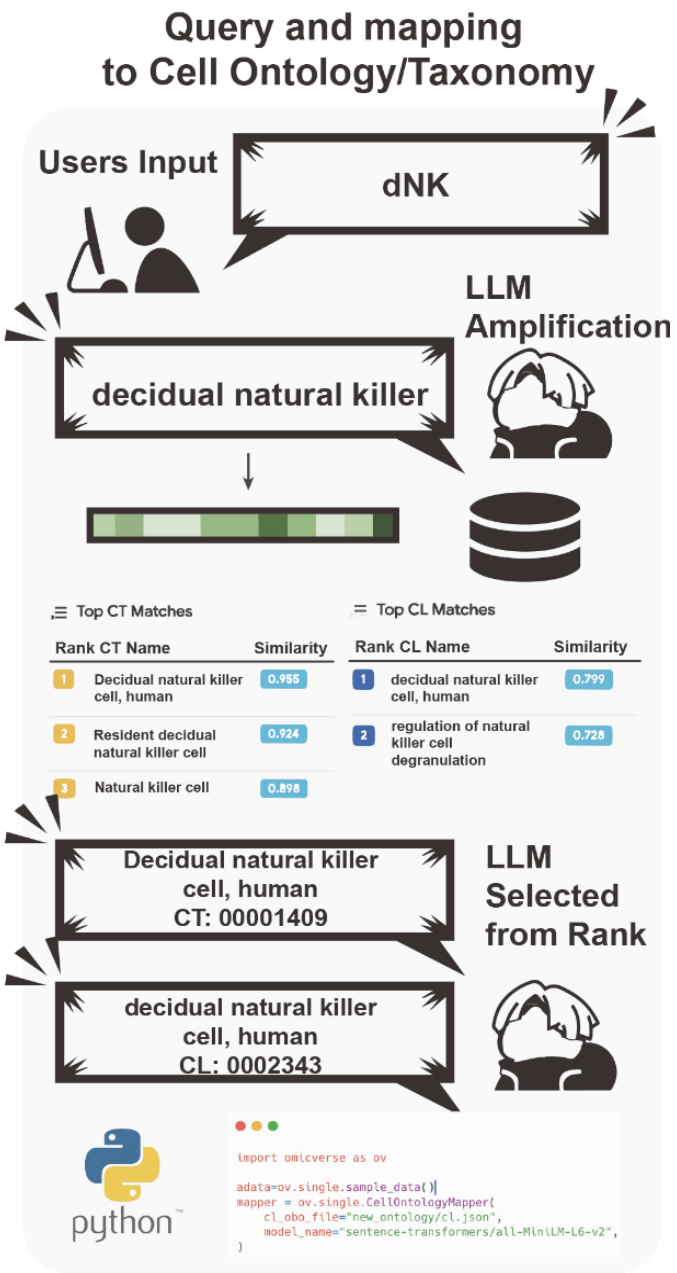
[3]斯坦福大学遗传学院 [4]卡罗林斯卡学院 [5]安徽医科大学公共卫生学院

Zehua Zeng, Xuehai Wang, Hongwu Du, Cencan Xing. 2025. CellOntologyMapper: Consensus mapping of cell type annotation. ***iMetaomics*** 2: e70064. https://doi.org/10.1002/imo2.70064

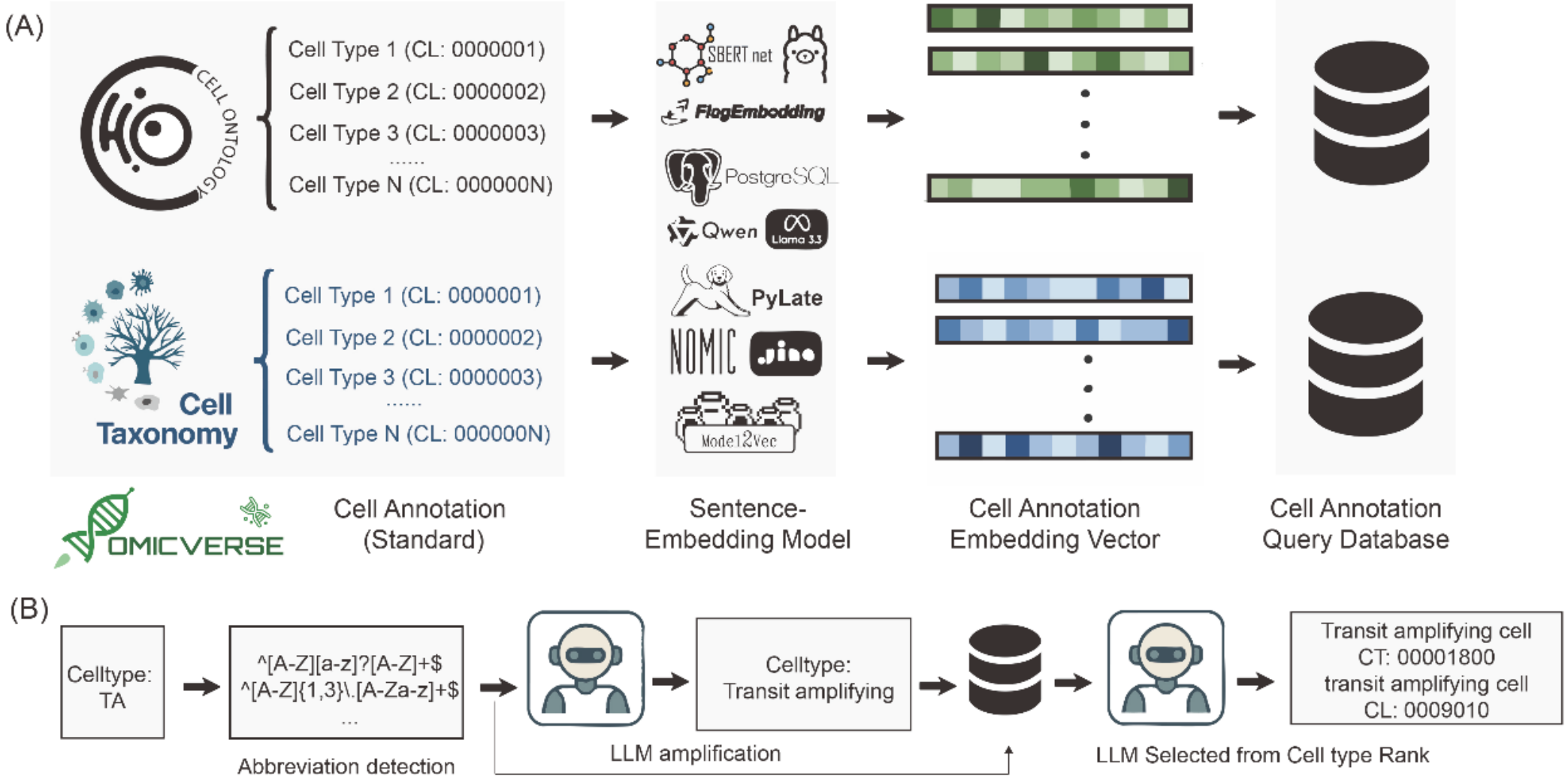
背景介绍

单细胞RNA测序的不同研究采用差异显著的细胞类型命名体系，导致数据整合与跨研究比较困难。例如，“NEUT”、“dM”或“Mac1”等名称可能指向相同或重叠的细胞群体，这一问题在跨物种研究中尤为突出。我们基于自然语言处理开发了一个自动化注释框架CellOntologyMapper，能够将用户自定义的细胞名称映射至标准化的本体标识，可有效处理缩写、同义词及组织来源等语境信息，并提供Python软件包和在线平台，便于集成至现有分析流程。





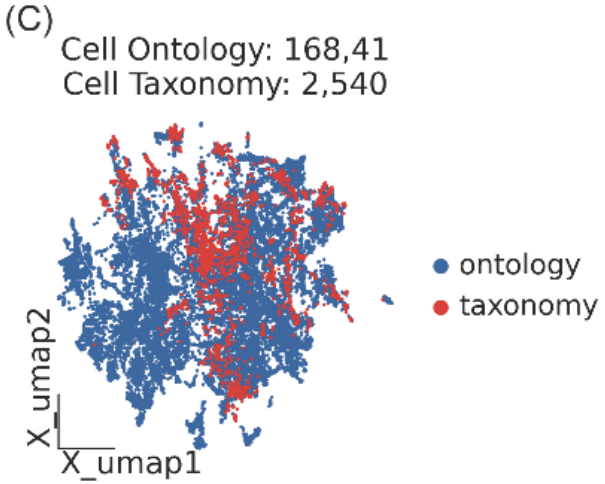
结果： CellOntologyMapper 设计概述

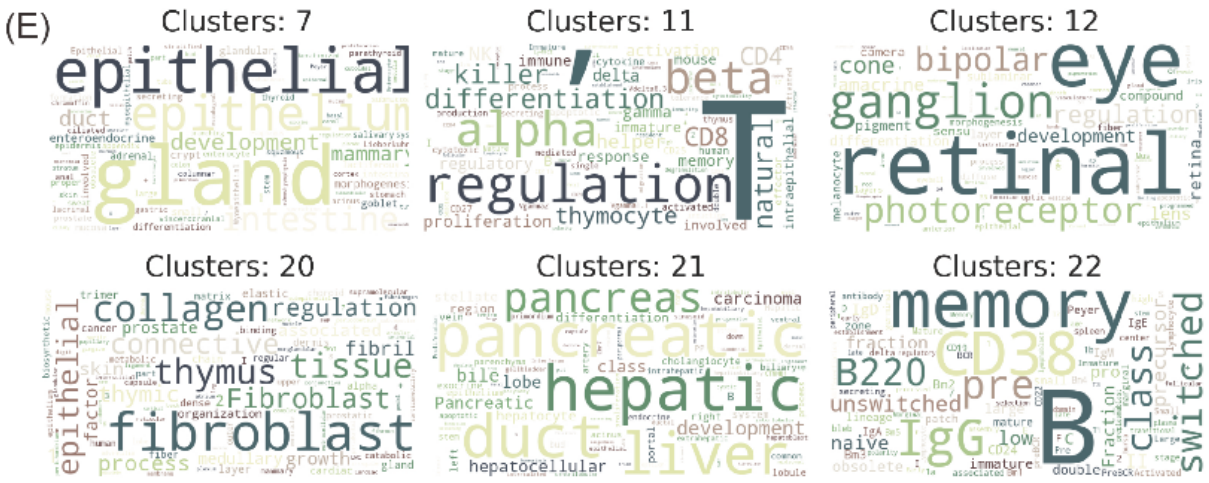


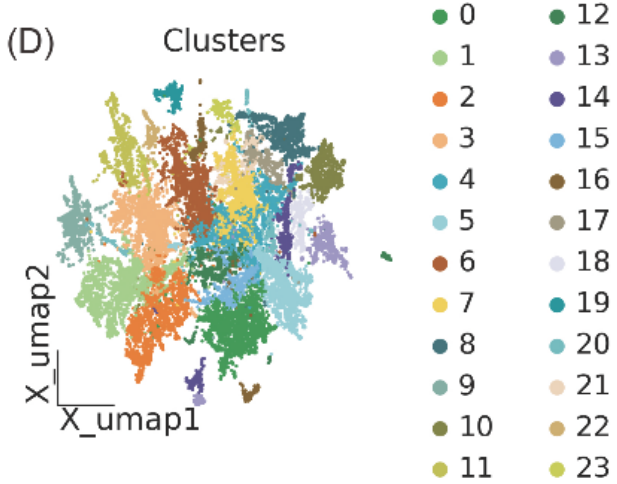
为了解决单细胞研究中细胞类型命名不一致的问题，CellOntologyMapper通过两个相互关联且互补的模块来标准化细胞类型命名：第一个模块构建了广泛的细胞本体查询数据库，作为准确细胞类型映射的基础资源（1A）；第二个模块实现了智能、具备语境感知能力的映射流程，可将用户提供的细胞类型术语转换为标准化的本体术语（1B）。

结果： CellOntologyMapper 设计概述

我们基于CT和CL构建的综合查询数据库包含 19,381 个精心整理的细胞类型条目（图 1C）。为研究该数据库的语义组织，我们使用 UMAP对术语嵌入向量进行降维，并应用无监督 Leiden 聚类，揭示出 24 个将语义相关细胞类型分组的独立簇（图 1D-E），其聚类模式验证了我们嵌入方法的生物学意义，并展示了该系统捕捉细胞分类学中细微关系的能力。







(F)

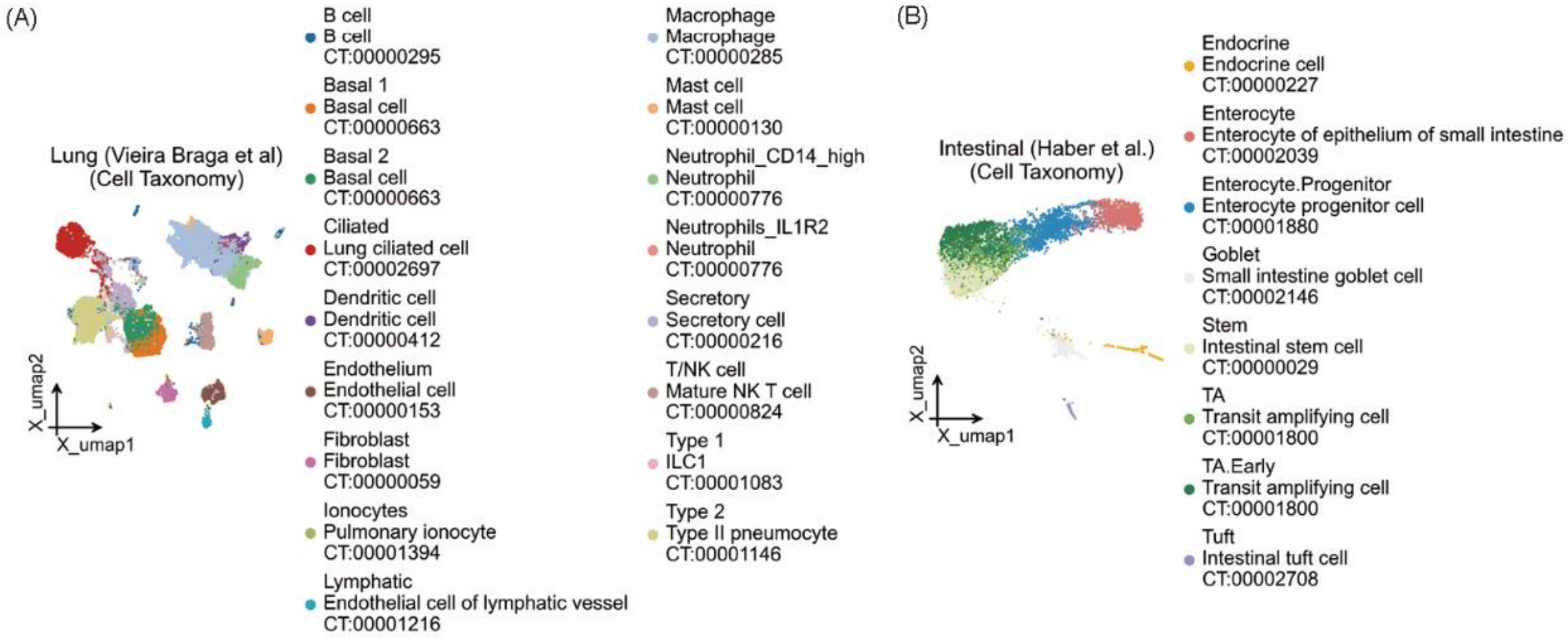
```python
import omicverse as ov

adata=ov.single.sample_data()
mapper = ov.single.CellOntologyMapper(
    cl_obo_file="new_ontology/cl.json",
    model_name="sentence-transformers/all-MiniLM-L6-v2",
)
results = mapper.map_adata_with_taxonomy(
    adata, cell_name_col="cell_label",new_col_name="enhanced_cell_ontology",
    species="Mus musculus",expand_abbreviations=True,use_taxonomy=True,
    tissue_context="Gut",threshold=0.5,
)
mapper.print_mapping_summary_taxonomy(results)
```

我们通过 Python 中的 OmicVerse 包提供了一个简洁、用户友好的应用程序接口（API），使研究人员能够用极少的代码完成细胞类型标准化（图 1F）。此外，我们还提供了一个在线网站(https://omicverse.com/cellmapper/）来帮助研究人员获得准确的细胞名称。

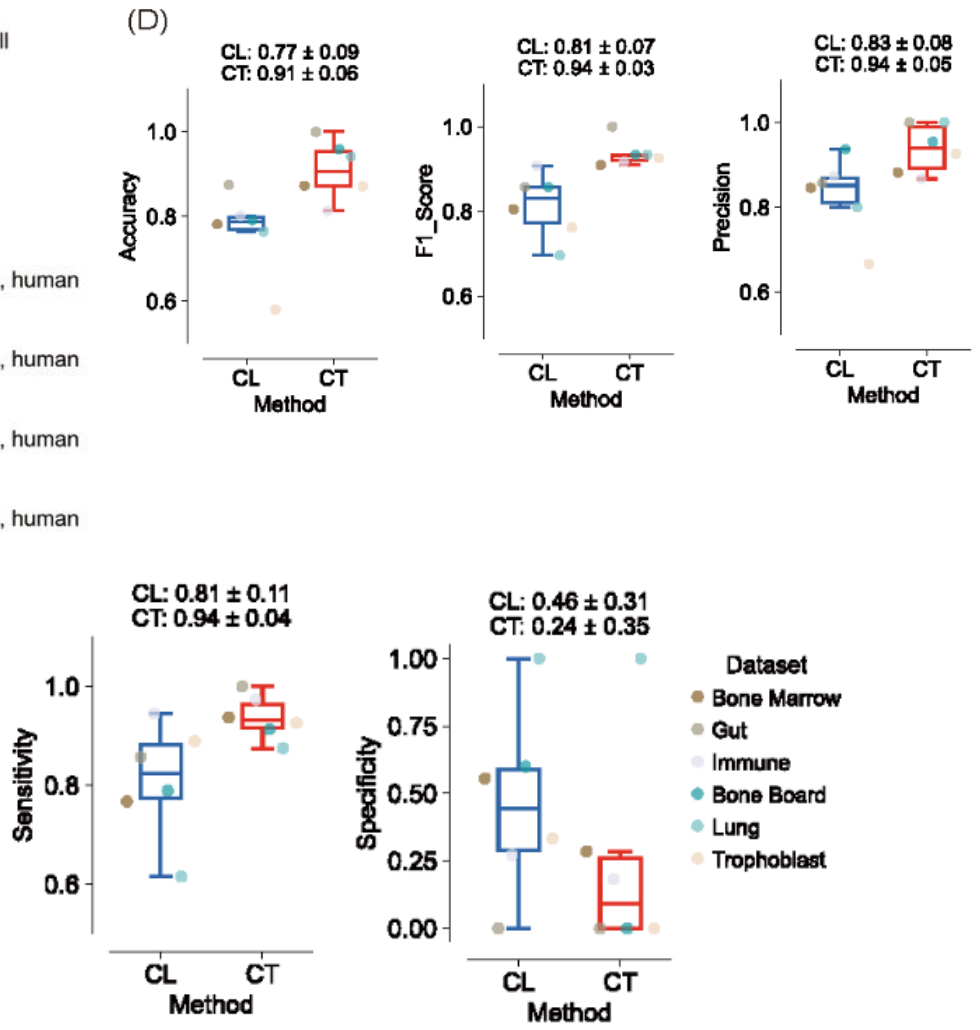
结果： CellOntologyMapper展现出稳健的跨尺度性能

为严格评估 CellOntologyMapper 的泛化性、准确性和稳定性，我们使用6个独立的单细胞数据集进行了全面验证，这些数据集覆盖了不同规模、生物学背景和注释复杂性。图2A为一个肺组织数据集的测试，图2B为一个肠道上皮数据集。CellOntologyMapper可准确注释过渡及多样化群体，并正确识别了如过渡扩增细胞和杯状细胞亚型等稀有群体，展现出其处理细胞身份跨越连续发育与分化谱系的复杂生物学环境的能力。



结果： CellOntologyMapper展现出稳健的跨尺度性能

胎盘滋养层细胞数据集更具挑战性，其包含多种胚胎细胞群体，且通常用缩写或简称表示。尽管存在这些重大挑战，CellOntologyMapper 仍成功解析了模糊缩写，并将每个注释群体准确映射到验证过的细胞分类学术语（图2C）。更为关键的是，该系统在不同发育轨迹间保持了生物学一致性，有效地区分了密切相关的祖细胞和分化滋养层状态，展示了其高级语义理解和语境注释的准确性。为进一步证实 CellOntologyMapper 的多样性和可扩展性，我们在5个不同规模的数据集上系统地评估了其准确率、F1 值、精确率、召回率和特异性（图2D）。



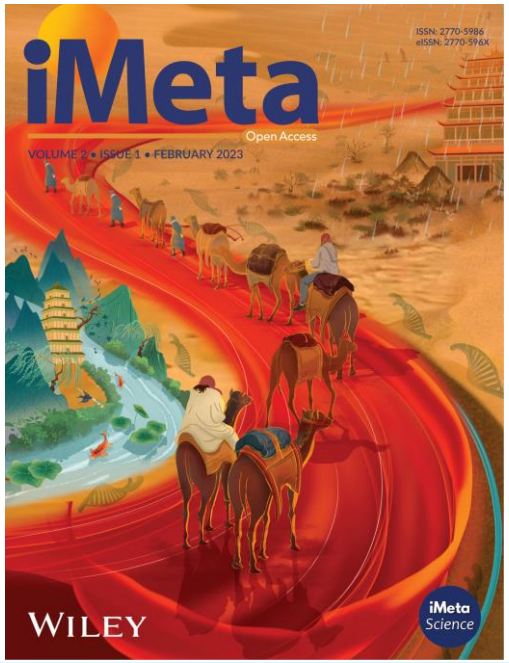
总结

- 利用先进的自然语言处理技术将模糊不清或简略的标签映射至**共识性的分类学框架**（细胞分类学和细胞分类）中，实现了细胞类型名称的标准化。
- 在不同规模和复杂度的数据集中均表现出**高映射准确度**（CL为0.835，CT为0.911），并能有效处理包含大量缩写和复杂发育谱系的挑战性案例。
- 提供 Python 包（独立仓库），模块（集成于 OmicVerse 生态系统）和用户友好的网页界面三两种形式，可**无缝集成**到现有单细胞分析流程，助力实现**可重复的大规模元分析。**
- 网址：https://omicverse.com/cellmapper/ 和 https://github.com/Starlitnightly/CellOntologyMapper

Zehua Zeng, Xuehai Wang, Hongwu Du, Cencan Xing. 2025. CellOntologyMapper: Consensus mapping of cell type annotation. ***iMetaomics*** 2: e70064. https://doi.org/10.1002/imo2.70064

iMeta(宏): 整合宏组学重新认识生物和医学

iMeta WILEY



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

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

主页: http://www.imeta.science

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

投稿: iMeta: https://wiley.atyponrex.com/journal/IMT2

iMetaOmics: https://wiley.atyponrex.com/journal/IMO2

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

office@imeta.science

imetaomics@imeta.science

宣传片

iMeta

更新日期 2025/7/6