CellOntologyMapper: Consensus mapping of cell type annotation

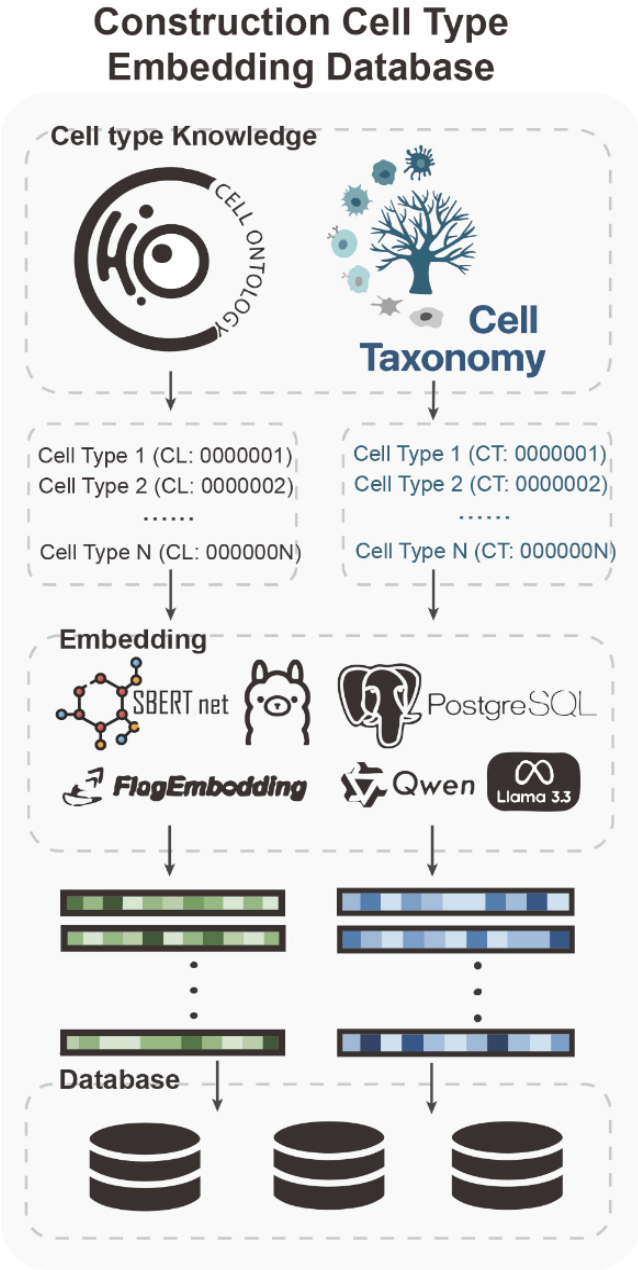
Zehua Zeng[1,2,3#], Xuehai Wang[4,5#], Hongwu Du[1,2*], Cencan Xing[1,2*]

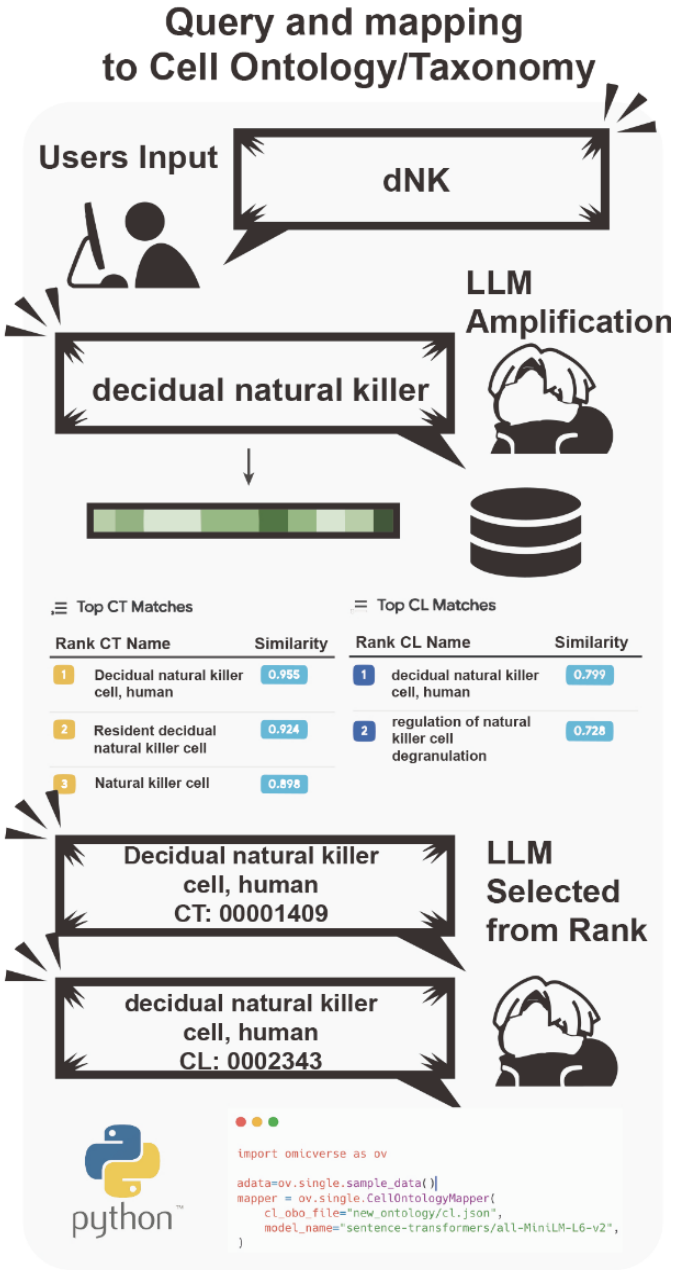
[1]1School of Chemistry and Biological Engineering, USTB, China.
[2]Daxing Research Institute, USTB, China.
[3]Department of Genetics, Stanford University, USA
[4]Karolinska Institute, Solna
[5]Department of Maternal, Child and Adolescent Health,
Anhui Medical University, Hefei, China.

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

Introduction

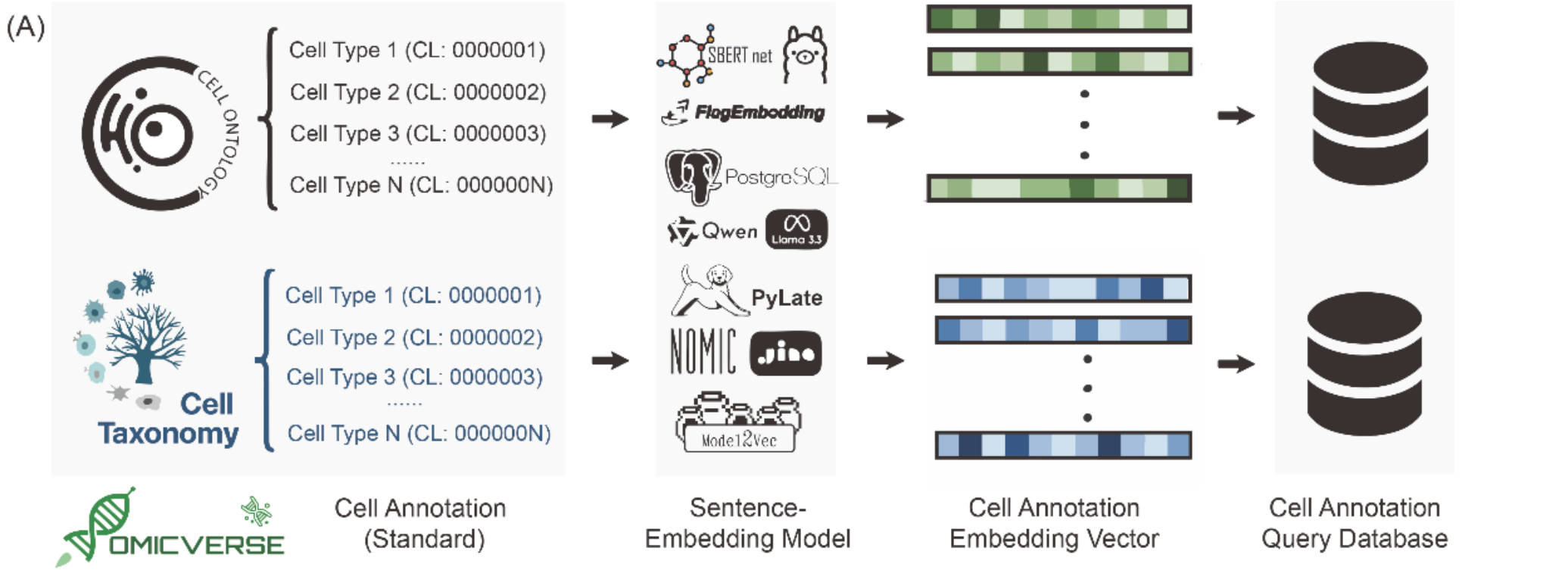
Single-cell RNA sequencing reveals cellular diversity, but inconsistent cell-type naming conventions—like "NEUT," "dM," or "Mac1"—hinder data integration and cross-study comparison, especially in cross-species research. We developed CellOntologyMapper, a computational tool that automatically maps custom cell names to standardized ontological terms using advanced natural language processing. It resolves abbreviations, synonyms, and tissue-specific contexts, and is available as a Python package in the OmicVerse, designed to integrate seamlessly into existing analysis workflows. Besides, user can also use our online website to perform the cell type name mapping (https://omicverse.com/cellmapper/).

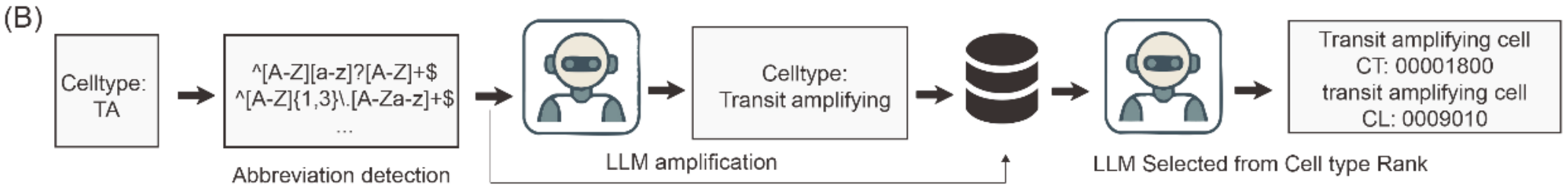




Results: Overview of CellOntologyMapper Design

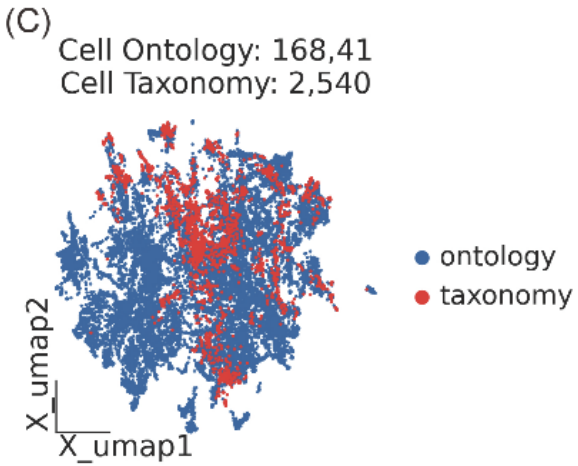
To address the challenge of inconsistent cell type naming, CellOntologyMapper standardizes cell type nomenclature through two interconnected and complementary modules: The first module constructs an extensive Cell Ontology Query Database, which serves as a foundational resource for accurate cell type mapping (1A); The second module implements an intelligent, context aware mapping pipeline aimed at converting user-provided cell type terms into standardized ontology terms (1B).

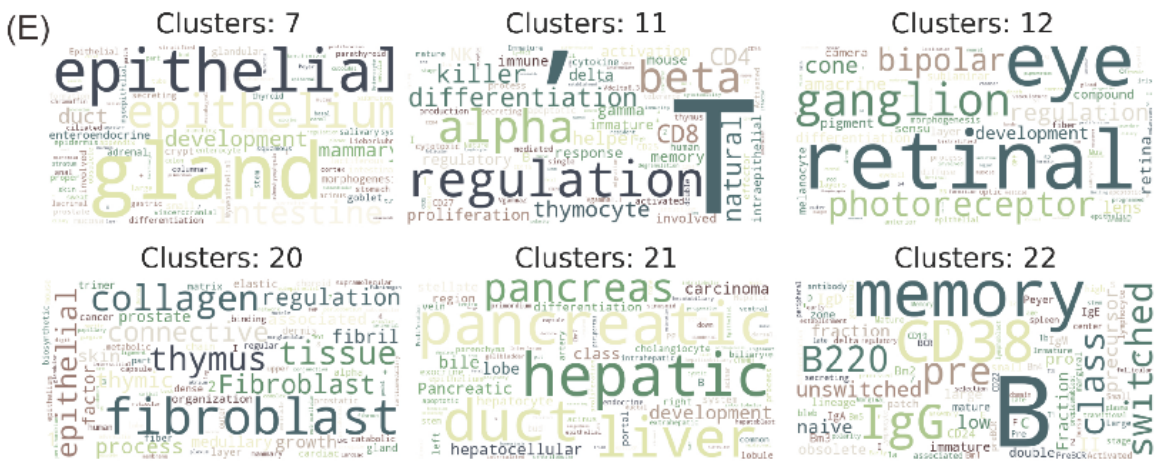


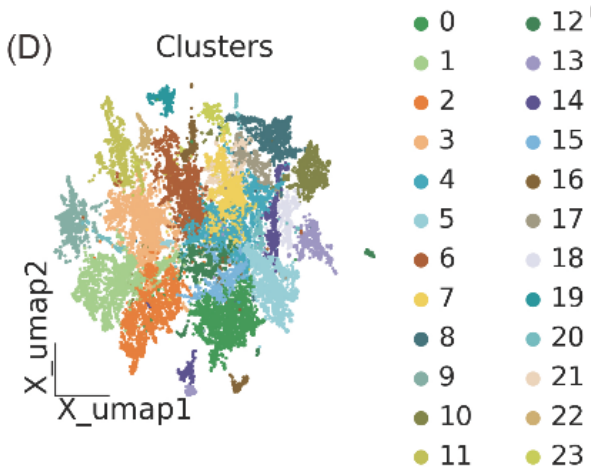


Results: Overview of CellOntologyMapper Design

The query database encompasses 19,381 curated cell type entries (1C). To investigate the semantic organization of this database, we performed dimensionality reduction using UMAP from the term embedding vector and applied unsupervised Leiden clustering, revealing 24 distinct clusters that group semantically related cell type (1D-E), validating the biological meaningfulness of our embedding approach and demonstrates the system's capacity to capture nuanced relationships within the cellular taxonomy.







(F)

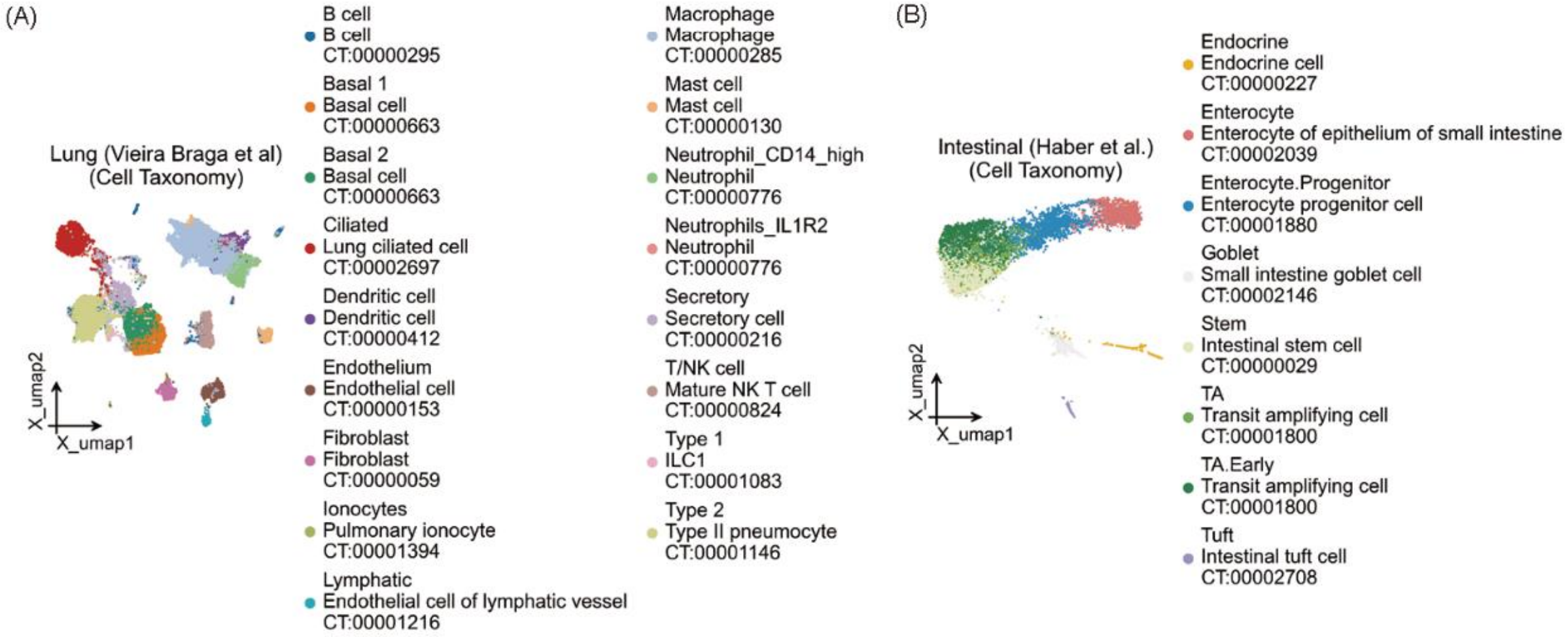
```python
import omicverse as ov

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)
```

We provide a user-friendly API through the OmicVerse Python package to perform cell type standardization (1F), and also provide an website https://omicverse.c-om/cellmapper/. This imple-mentation facilitates seam-less integration into existing single-cell analysis work-flows while maintaining the sophisticated mapping cap-abilities of the underlying framework.

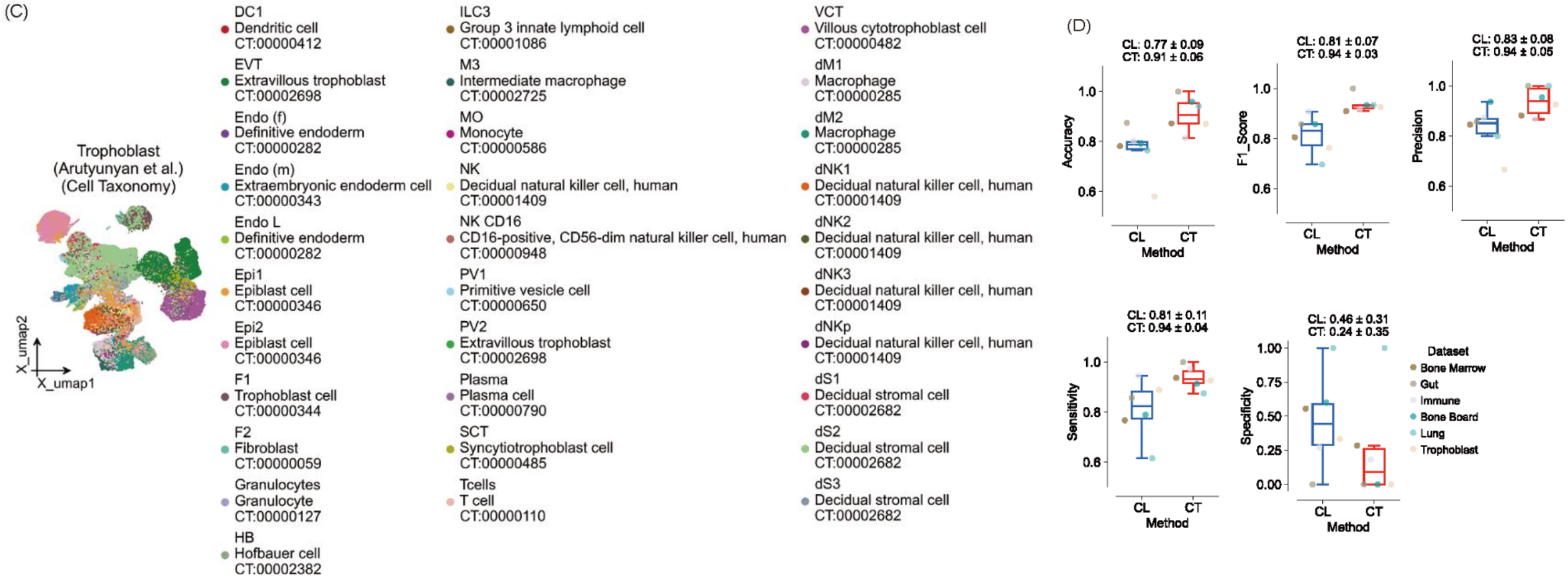
Results: The robust cross-scale performance

To rigorously evaluate the versatility, precision and robustness, we conducted comprehensive validation using multiple independent single cell datasets spanning diverse scales, biological contexts and annotation complexities. CellOntologyMapper can accurately annotate transitional and diverse populations, and correctly identified rare populations like Transit amplifying and goblet cell subtypes, demonstrating its ability to handle nuanced biological contexts in which cell identity spans continuous developmental and differentiation spectra.



Results: The robust cross-scale performance

We subjected CellOntologyMapper to an even more stringent test using a highly complex trophoblast dataset, incorporated various embryonic cell populations and often represented by acronyms and short forms common. Despite these substantial challenges, CellOntologyMapper exhibited outstanding performance, successfully resolving ambiguous abbreviations and accurately mapping each annotated population to validated Cell Taxonomy terms. Critically, the system preserved biological coherence across distinct developmental trajectories, effectively differentiating closely related progenitor and differentiated trophoblast states, demonstrating sophisticated semantic understanding and contextual annotation accuracy.



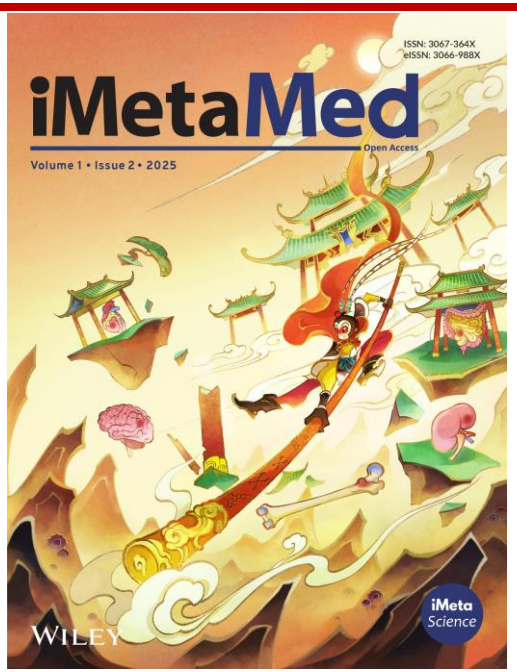
Summary

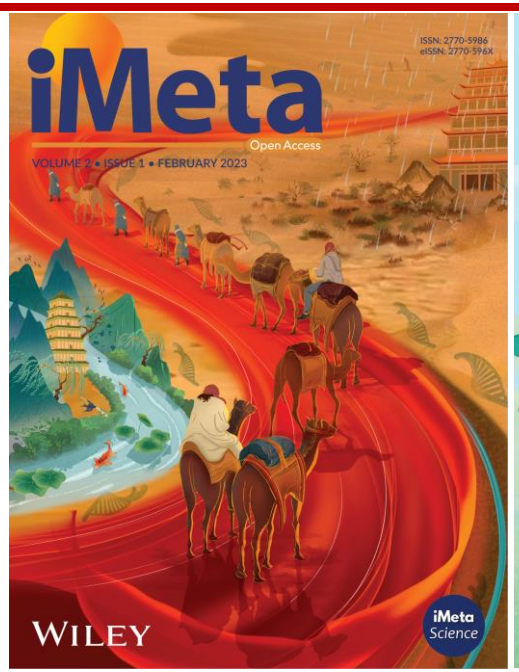
- CellOntologyMapper automatically maps diverse and often ambiguous user-defined cell type names to standardized terms from the Cell Ontology and Cell Taxonomy using advanced NLP techniques, including sentence transformers and large language models.
- The tool achieves high mapping accuracy (0.835 for CL, 0.911 for CT) across datasets of varying scales and complexities, including challenging cases with extensive abbreviations and developmental lineages.
- Available as both a Python package within the OmicVerse ecosystem and a user-friendly web interface, facilitates seamless integration into existing single-cell analysis workflows, promoting reproducible and large-scale meta-analyses.
- Website: https://omicverse.com/cellmapper/ and 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: To be top journals in biology and medicine

WILEY





"***iMeta***" launched in 2022 by iMeta Science Society, impact factor (IF) **33.2**, ranking top 65/22249 in world and 2/161 in the microbiology. It aims to publish innovative and high-quality papers with broad and diverse audiences. **Its scope is similar to *Cell, Nature Biotechnology/Methods/Microbiology/Medicine/Food***. Its unique features include video abstract, bilingual publication, and social media with 600,000 followers. Indexed by **SCIE/ESI**, **PubMed**, **Google Scholar** etc.

"***iMetaOmics***" launched in 2024, with a **target IF>10, and its scope is similar to *Nature Communications*, *Cell Reports***, *Microbiome, ISME J, Nucleic Acids Research, Briefings in Bioinformatics,* etc.

"***iMetaMed***" launched in 2025, with a **target IF>15, similar to *Med, Cell Reports Medicine, eBioMedicine, eClinicalMedicine*** etc.

Society: http://www.imeta.science
Publisher: https://wileyonlinelibrary.com/journal/imeta
iMeta: https://wiley.atyponrex.com/journal/IMT2
Submission: iMetaOmics: https://wiley.atyponrex.com/journal/IMO2
iMetaMed: https://wiley.atyponrex.com/journal/IMM3

iMetaScience
iMetaScience

office@imeta.science
imetaomics@imeta.science

Promotion Video

Update 2025/7/6