



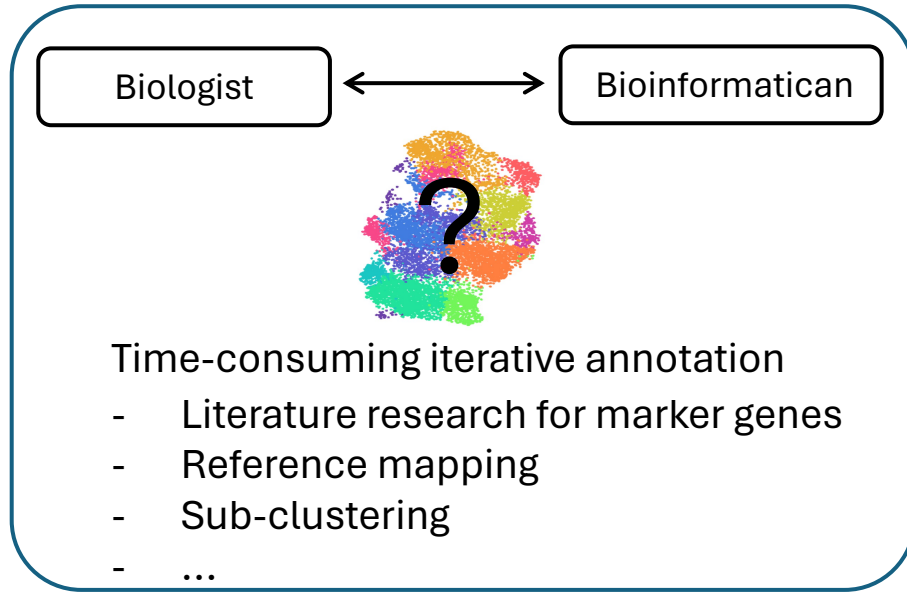
You can talk to cells now!

Joint Embedding of Transcriptomes and Text Enables Interactive Single-Cell RNA-seq Data Exploration via Natural Language

Moritz Schaefer^{1,2}, Peter Peneder^{3,4}, Daniel Malz^{2,5}, Anna Hakobyan^{2,5}, Varun Sharma^{1,2,5}, Thomas Krausgruber^{1,2}, Jörg Menche^{2,5,6}, Eleni M. Tomazou³, Christoph Bock^{1,2}

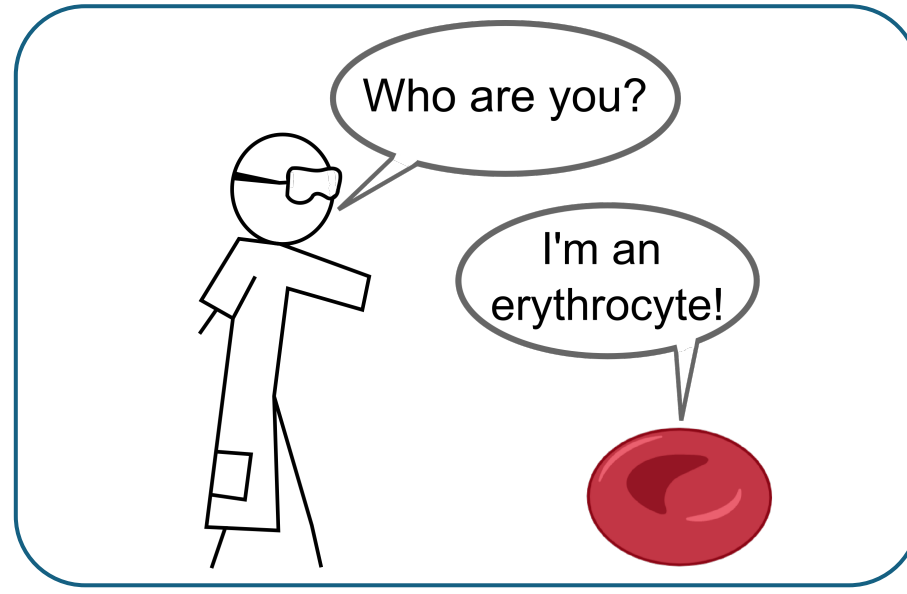
Problem

Understanding cells is hard



Solution

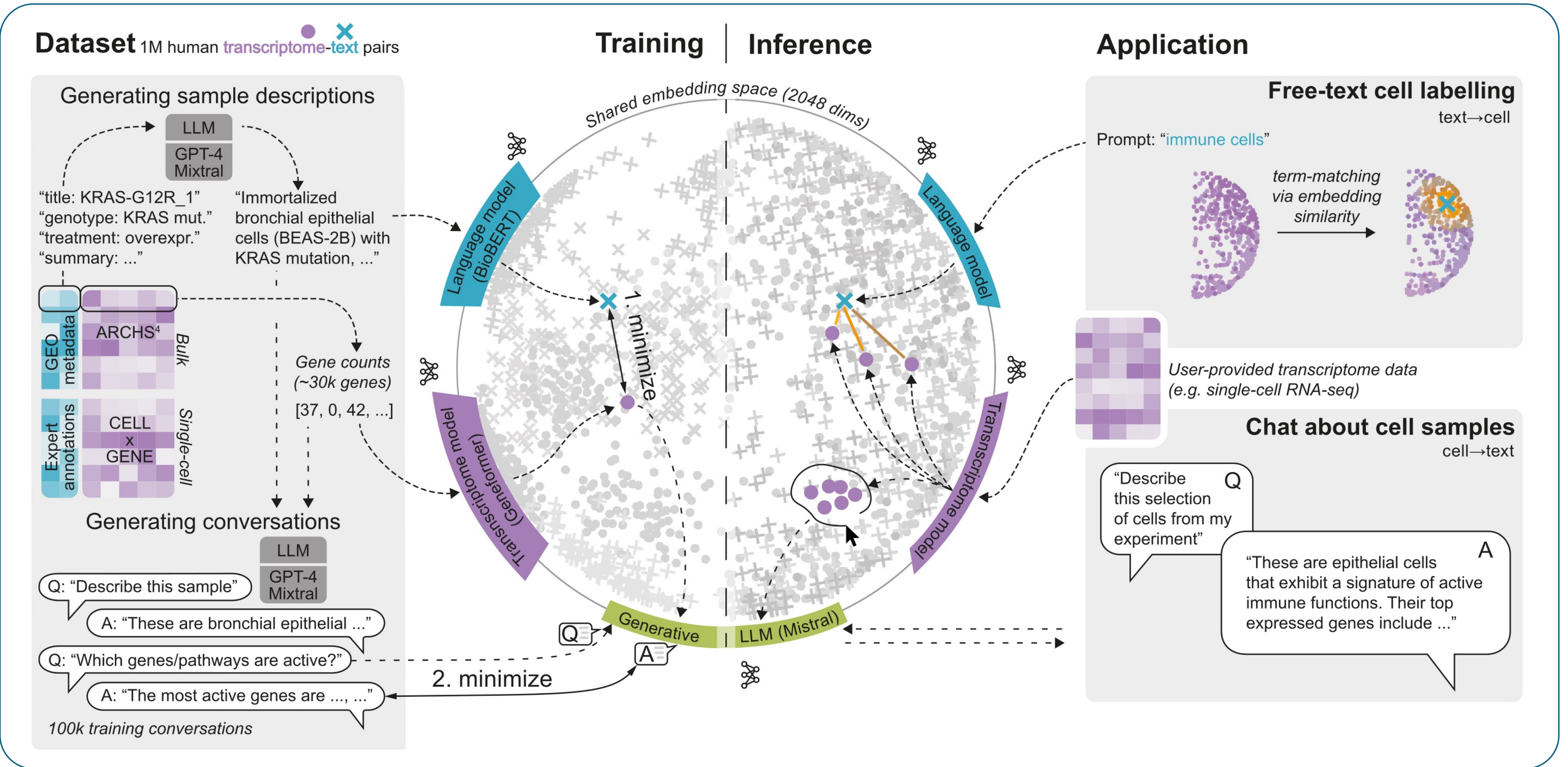
Just talk to them



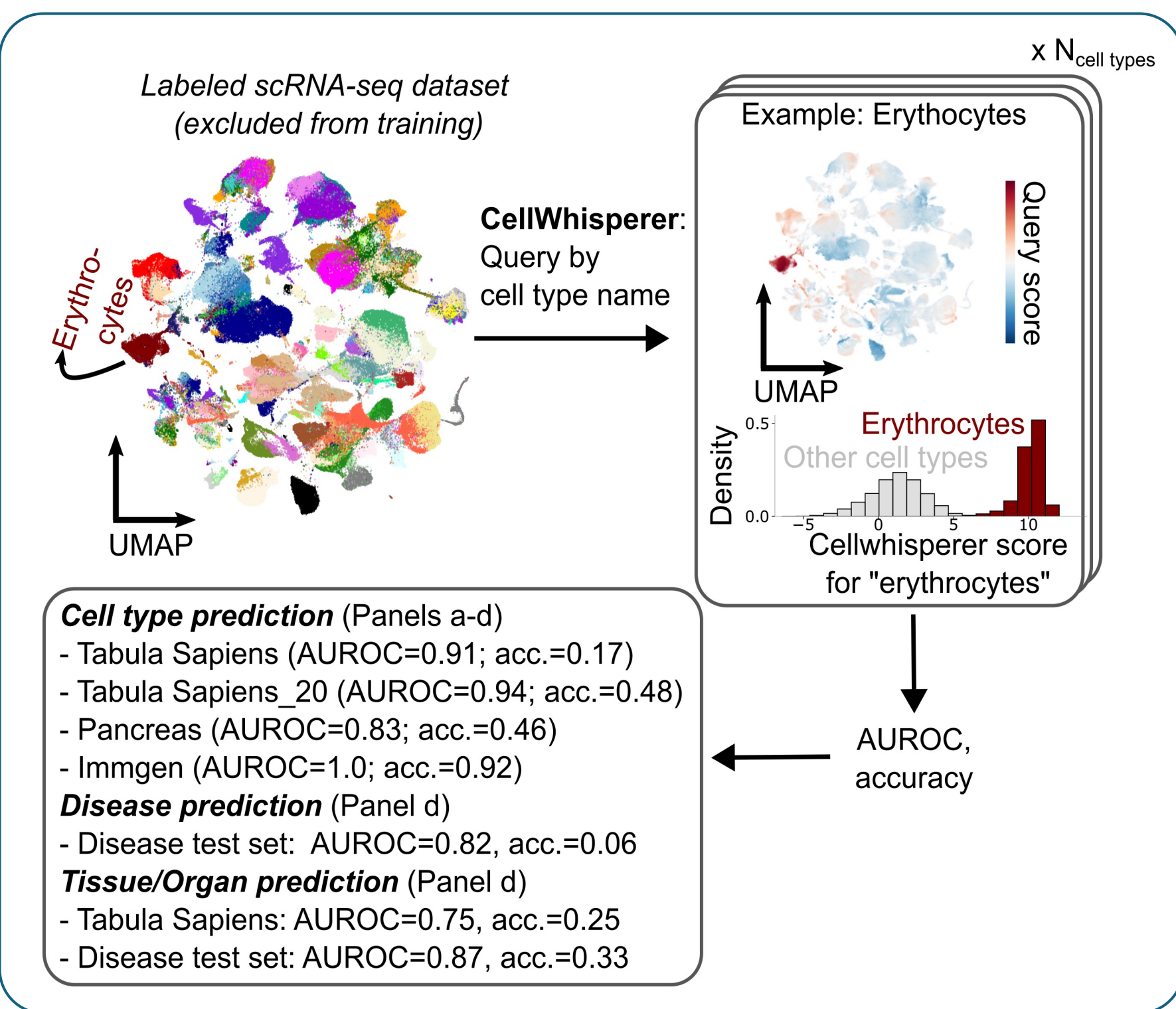
Abstract

Single-cell RNA sequencing (scRNA-seq) has revolutionized our understanding of cellular states, but interpreting the vast data it generates remains challenging. Here, we introduce CellWhisperer, a multimodal machine learning model that bridges the gap between transcriptomics data and natural language, enabling intuitive interaction with scRNA-seq datasets. Trained on one million pairs of transcriptomes and their textual annotations from the Gene Expression Omnibus (bulk) and CELLxGENE census (single-cell), CellWhisperer employs contrastive learning to create a joint embedding space, enabling tasks such as cell retrieval based on free-text queries and zero-shot classification of a broad range of properties such as cell types, tissue of origin, and disease context. Based on this space, we fine-tune a generative large-language model to answer questions about selected transcriptomes via a chat interface. Integrated into the CELLxGENE browser, CellWhisperer allows biologists to explore single-cell transcriptomes using natural language queries. Our experiments show that CellWhisperer accurately annotates cellular states, without relying on reference datasets. This work paves the way for easier and better interpretation of scRNA-seq data, including those that are poorly covered by reference datasets, leveraging the power of natural language for transcriptomics research.

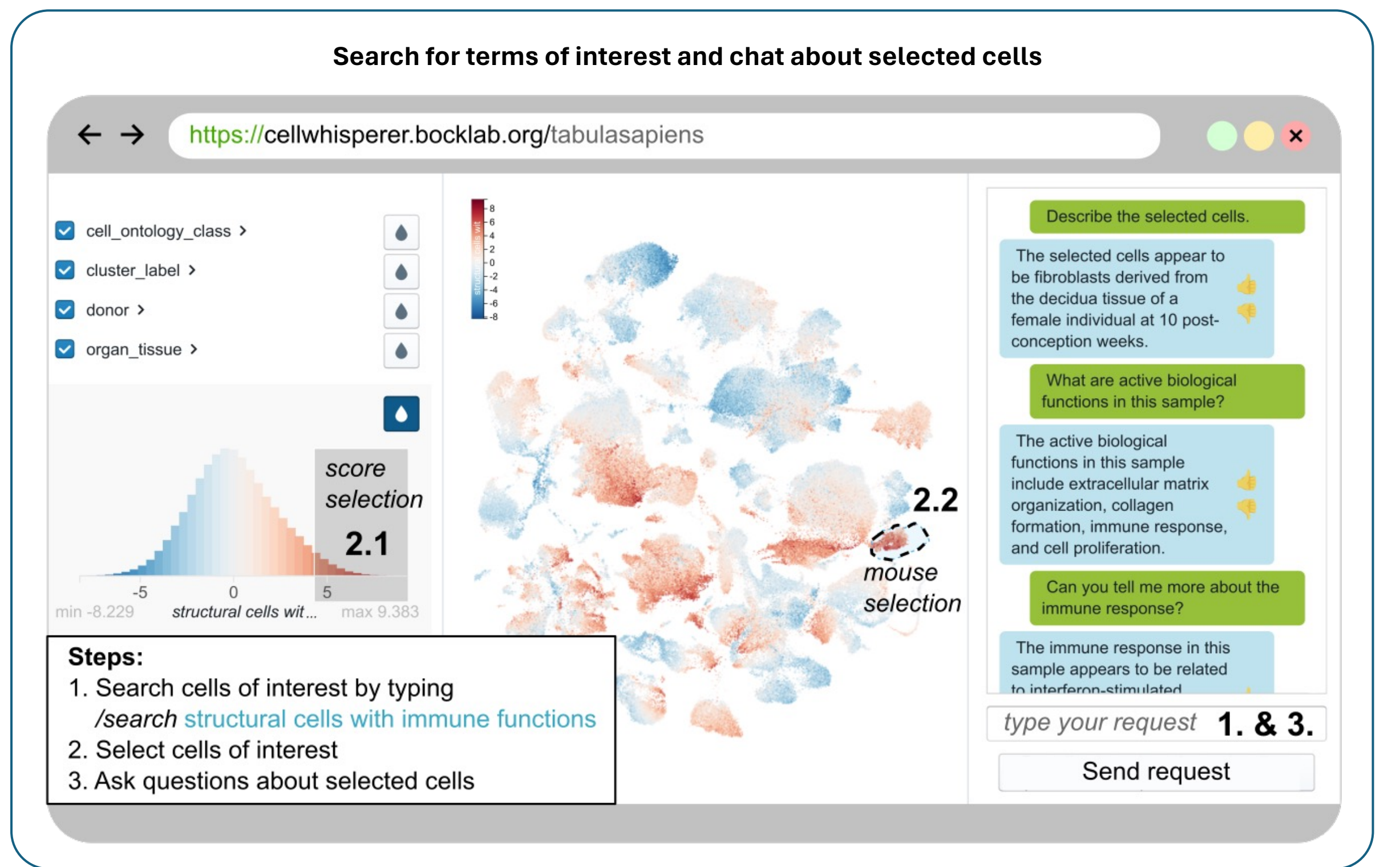
Methodology



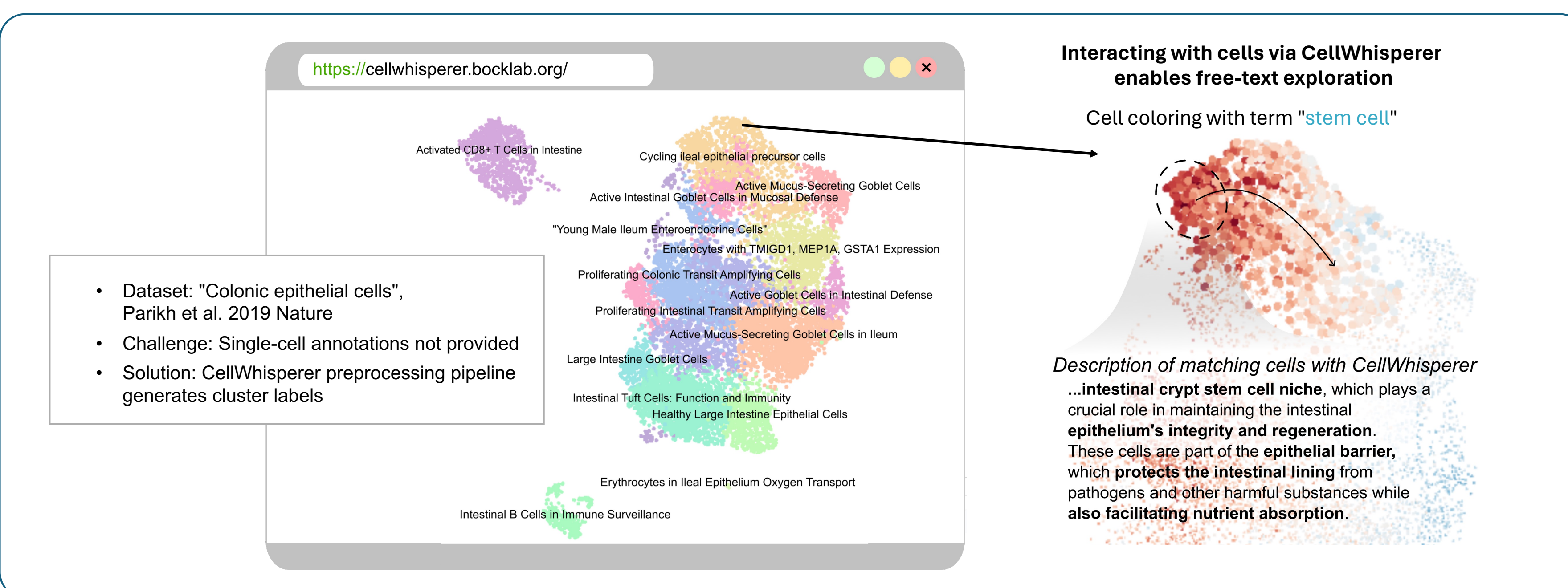
Validation



Inference



Example



5. Follow us!

Preprint&Code available soon!

@muronglizi
@BockLab

@moritzbaio
.bsky.social

Test it now:

<https://cellwhisperer.bocklab.org>

- Affiliations:**
1. Institute of Artificial Intelligence, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria
 2. CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria
 3. St. Anna Children’s Cancer Research Institute (CCRI), Vienna, Austria
 4. Doctoral School in Microbiology and Environmental Science, University of Vienna, Vienna, Austria.
 5. Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, Vienna, Austria
 6. Faculty of Mathematics, University of Vienna, Vienna, Austria
- * Equal contribution