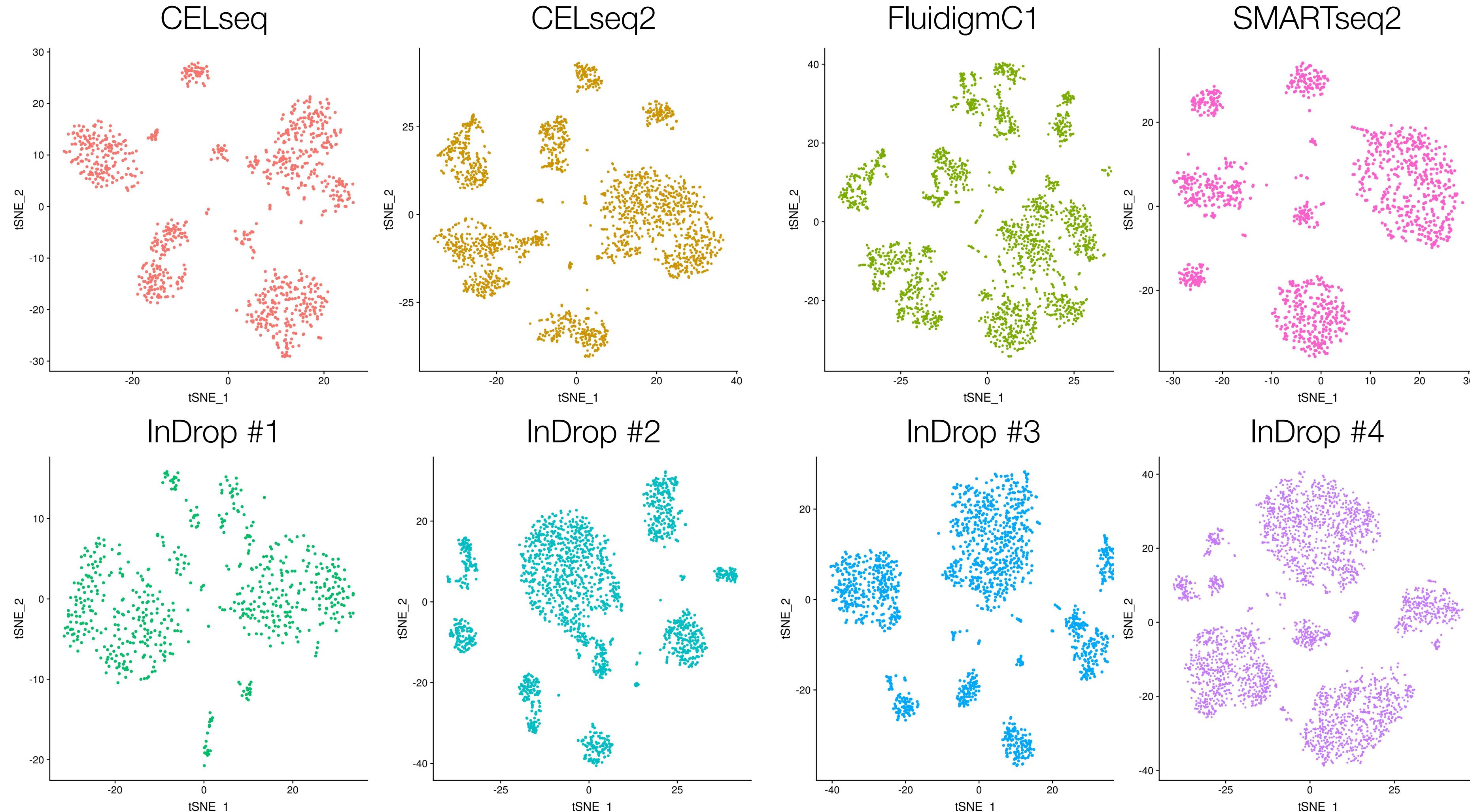


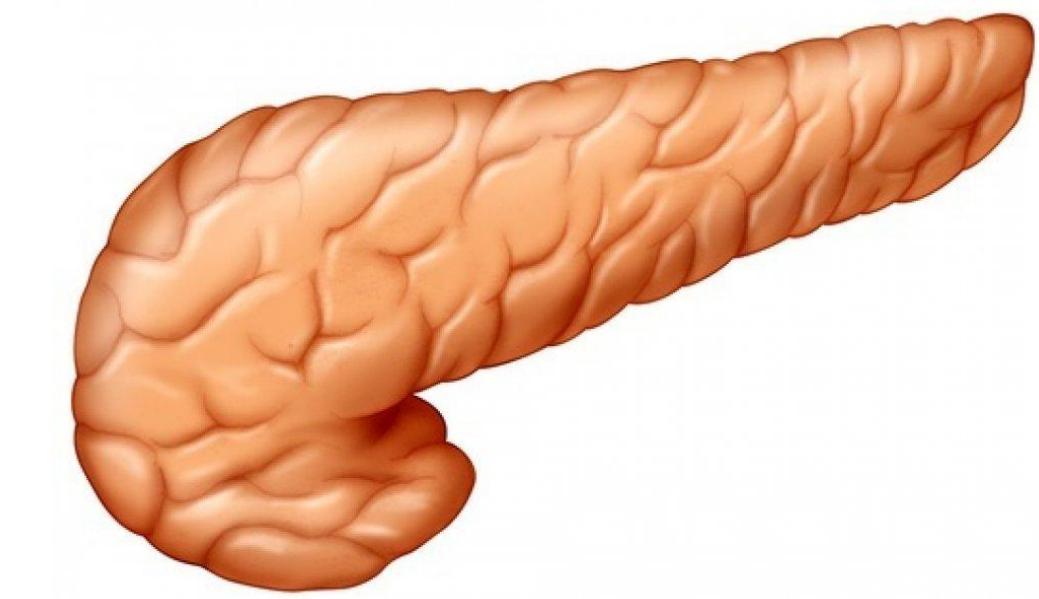
Seurat v3 integration

Building a cell atlas

8 maps of the human pancreas



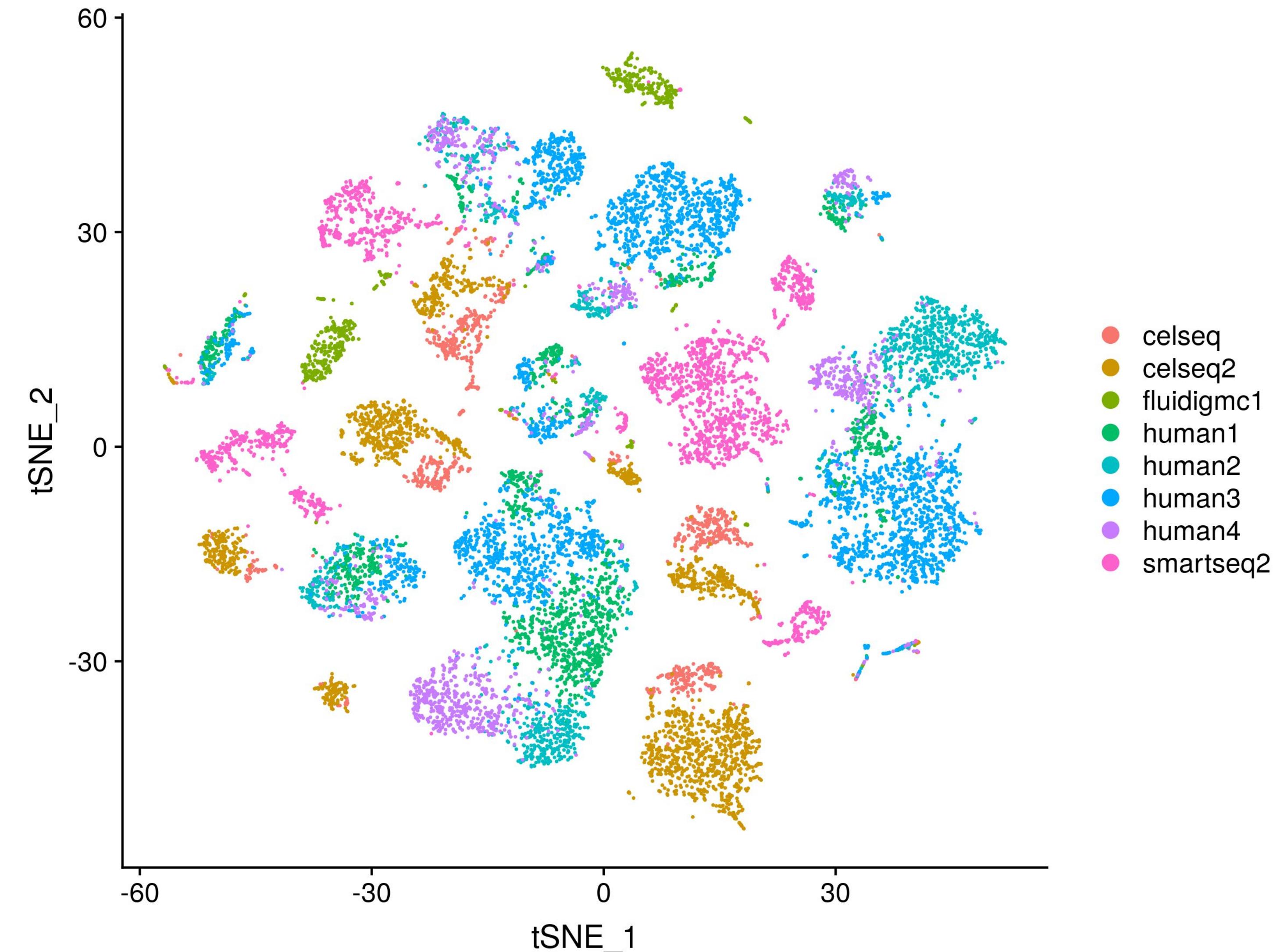
Human pancreas



Baron et al. 2016, *Cell Syst.*
Lawlor et al. 2017, *Genome Res.*
Grun et al. 2016, *Cell Stem Cell*
Muraro et al. 2016, *Cell Syst.*

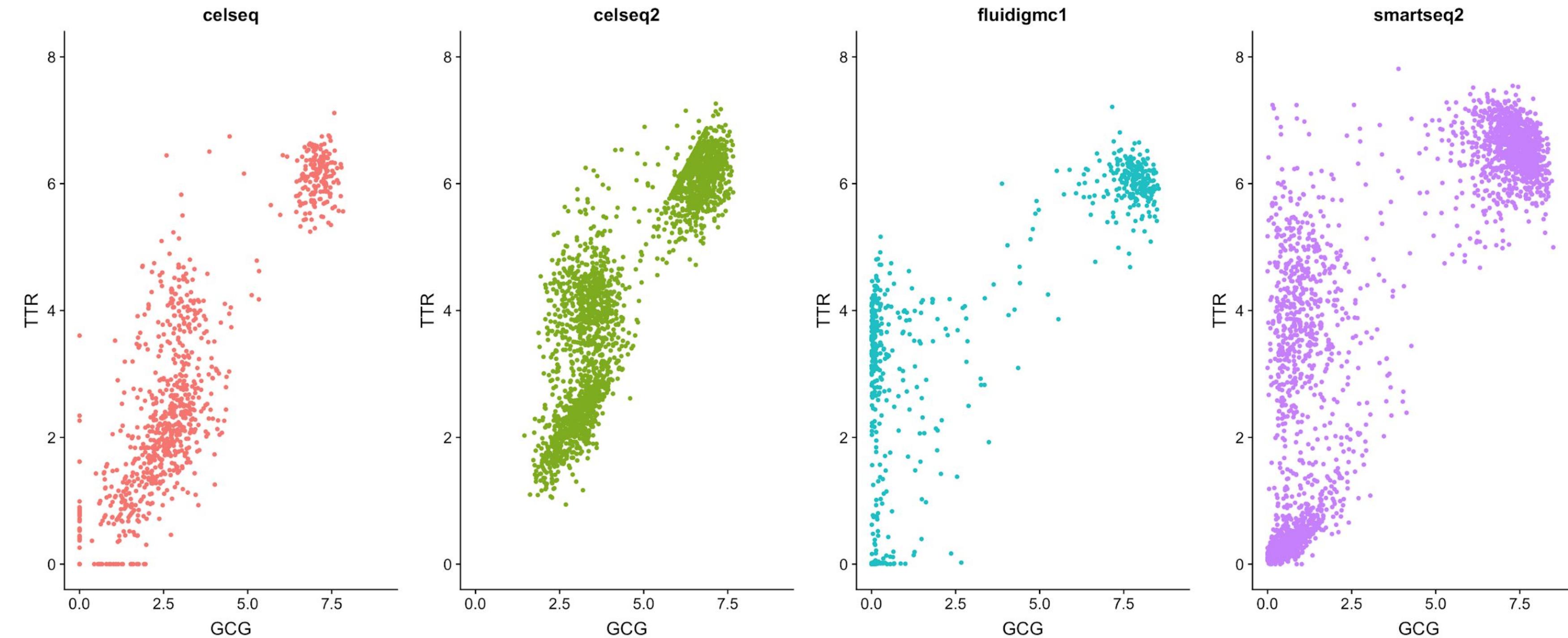
Building a cell atlas

8 maps of the human pancreas



How can we jointly analyze
datasets in the presence of
significant batch variation?

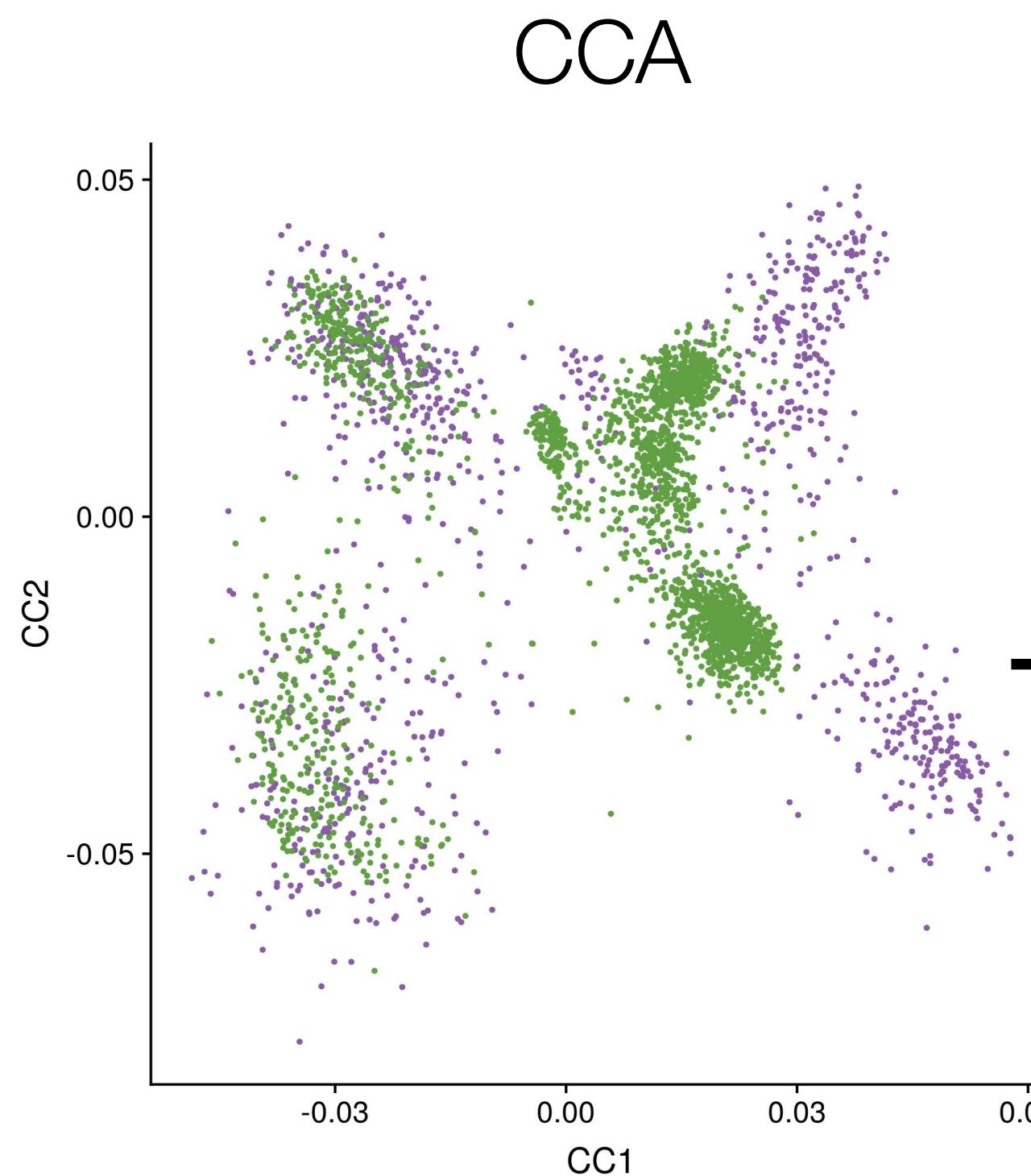
Integrated analysis across experiments



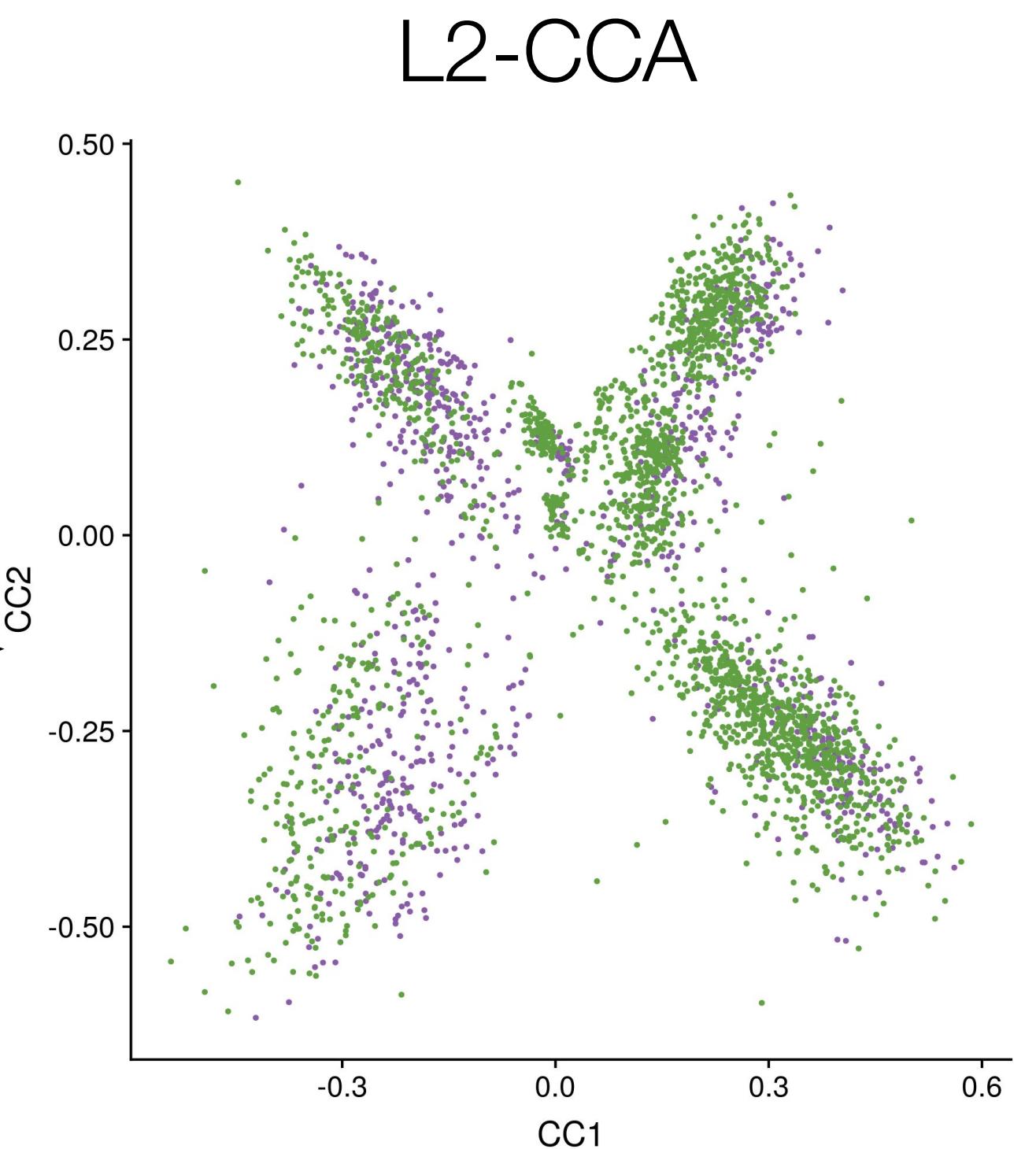
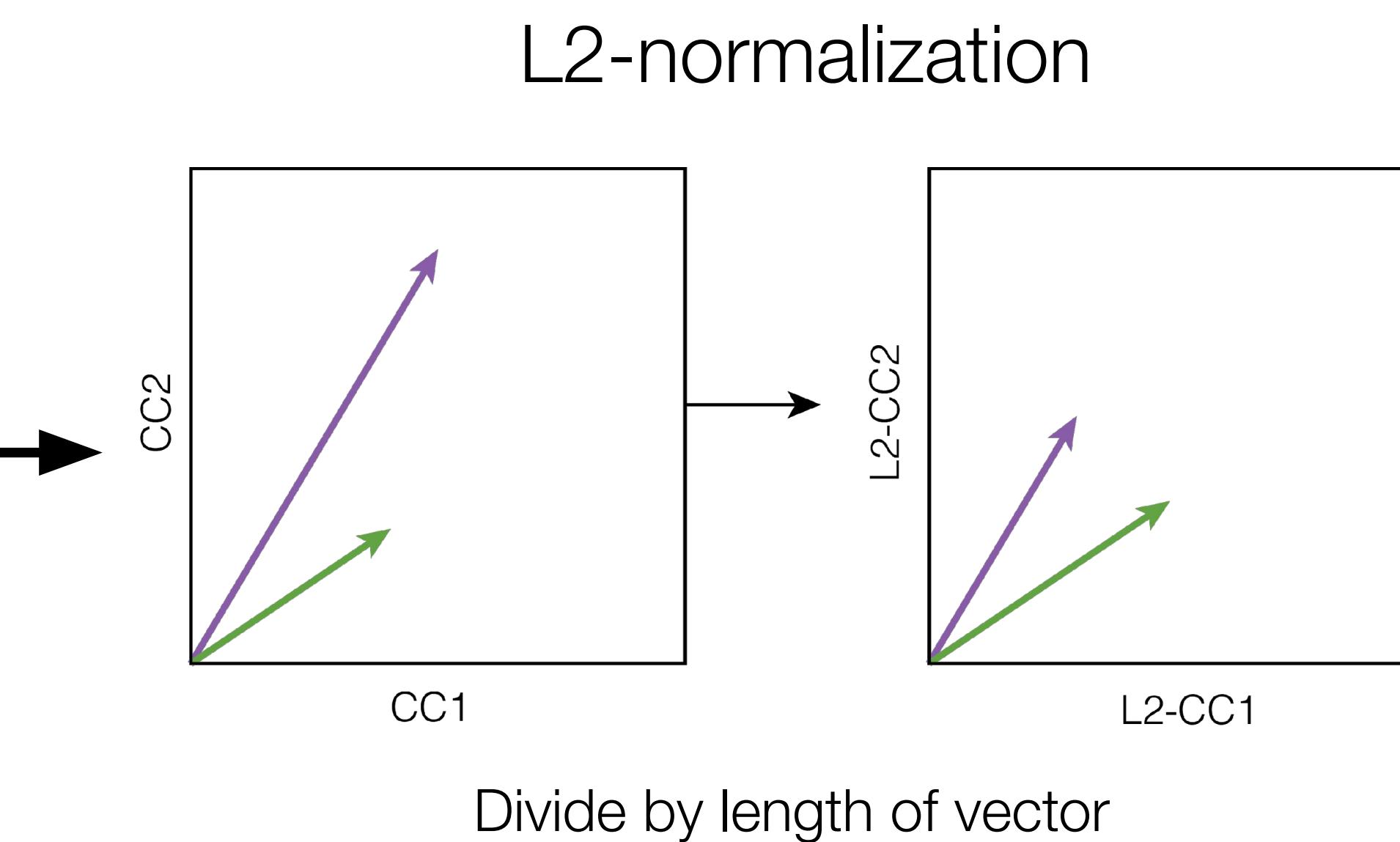
Intuition: identify **shared sources of variation** across datasets to identify equivalent cell states

Finding corresponding cells

Canonical correlation analysis and normalization



Butler et al. 2018, *Nature Biotechnology*

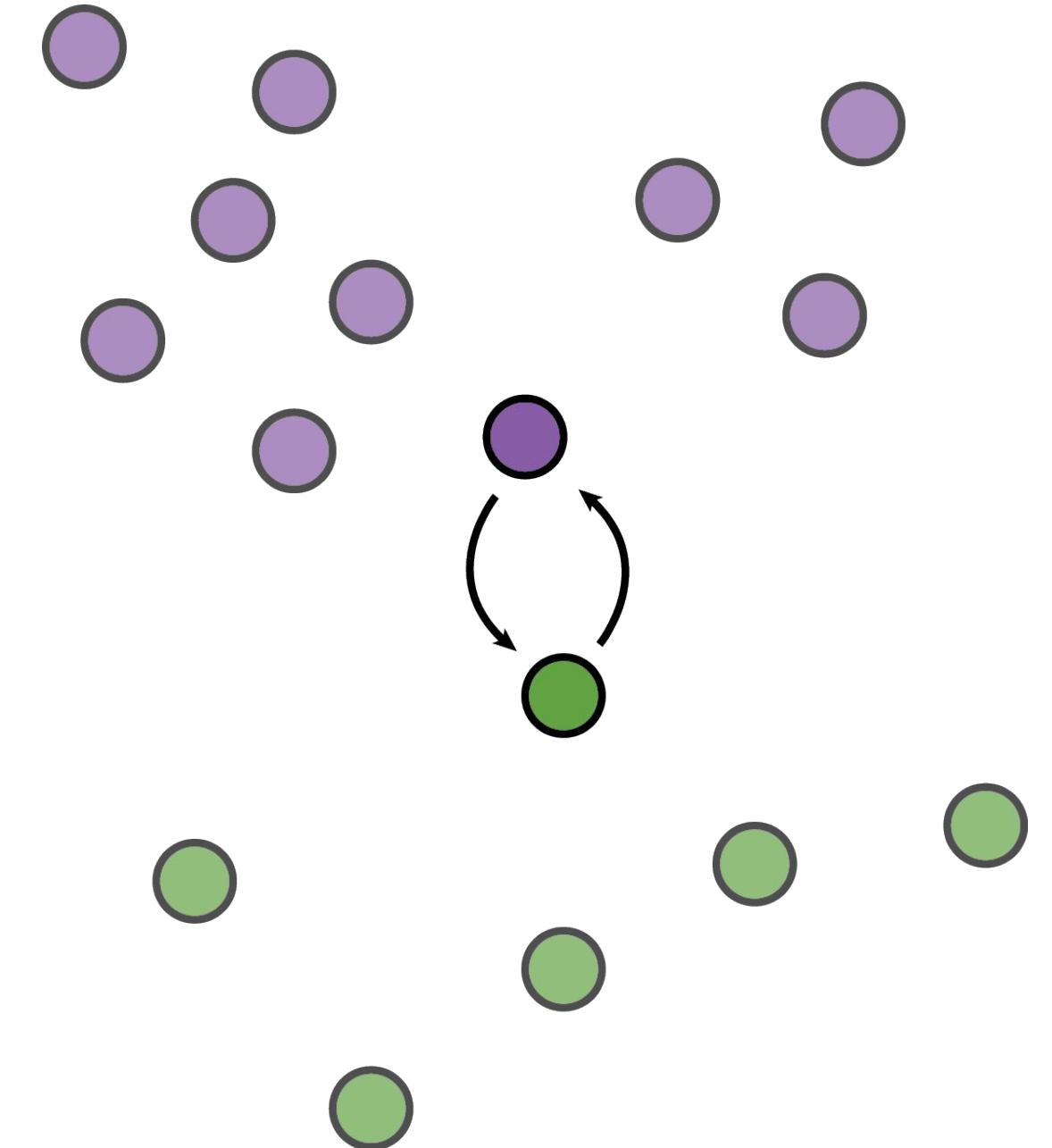


CCA captures correlated sources of variation between two datasets

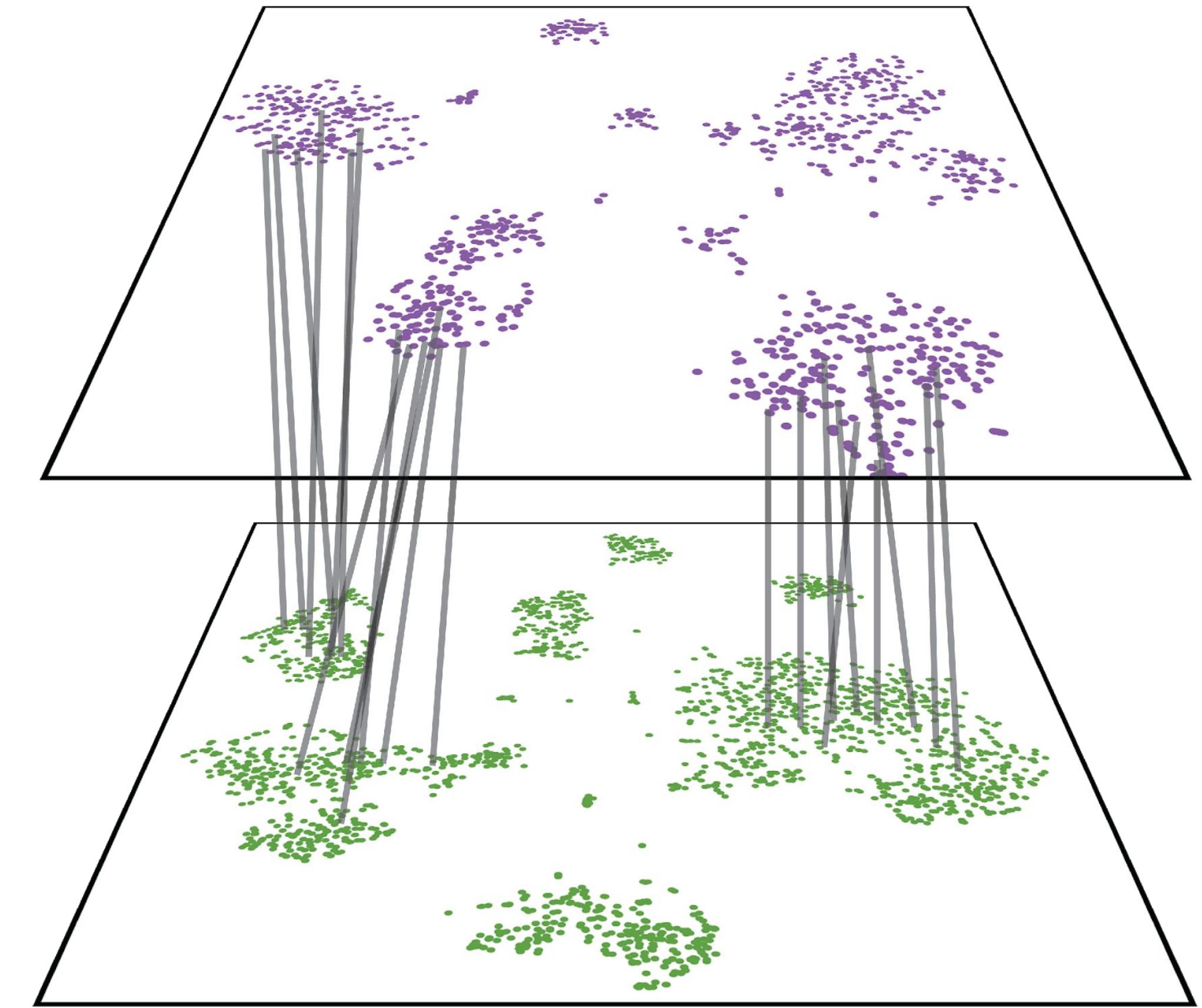
L2-normalization corrects for differences in scale

Finding corresponding cells

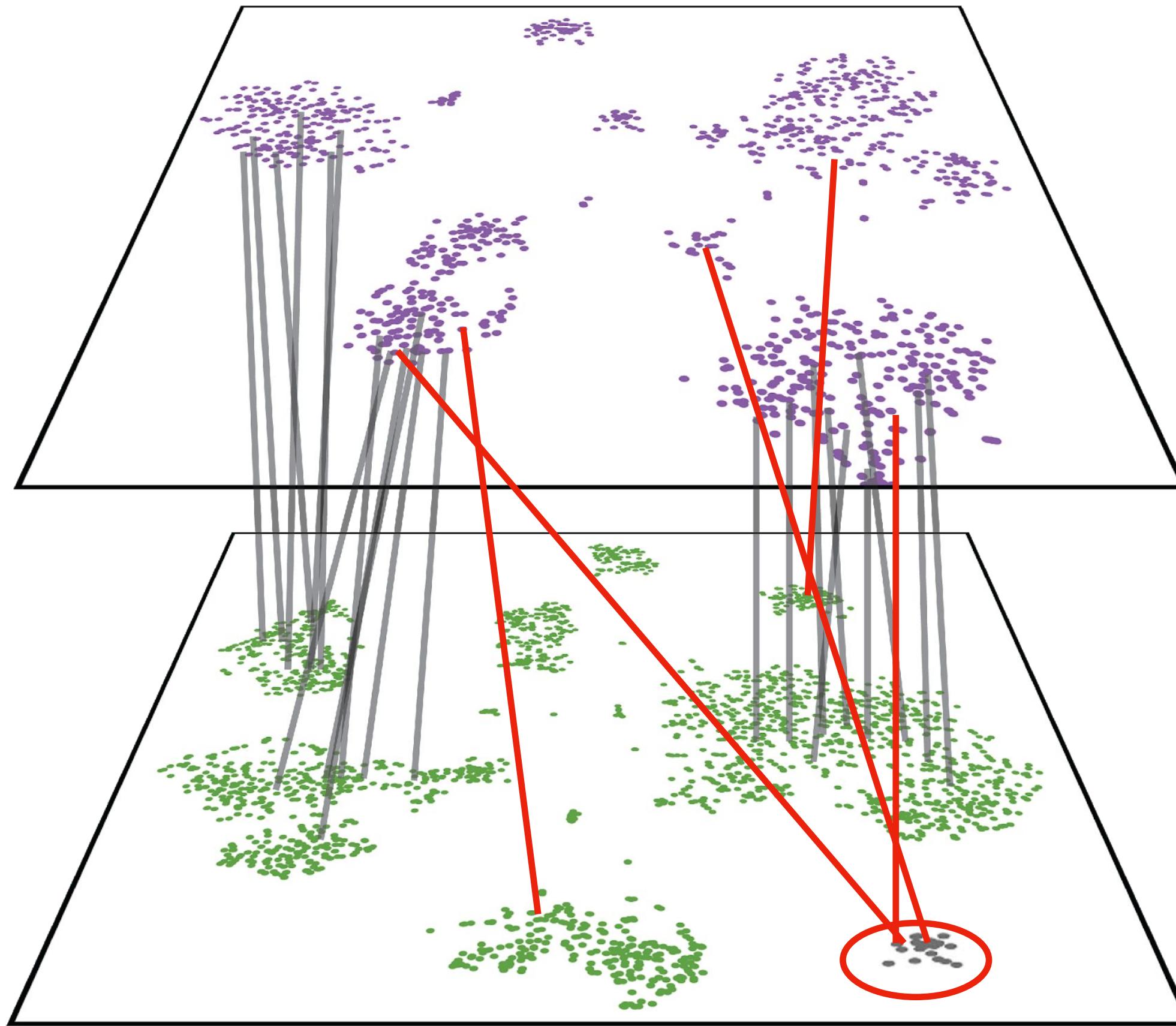
Mutual nearest neighbors



Cells that are **mutually** close to one another are equivalent
We call these cells **anchors**

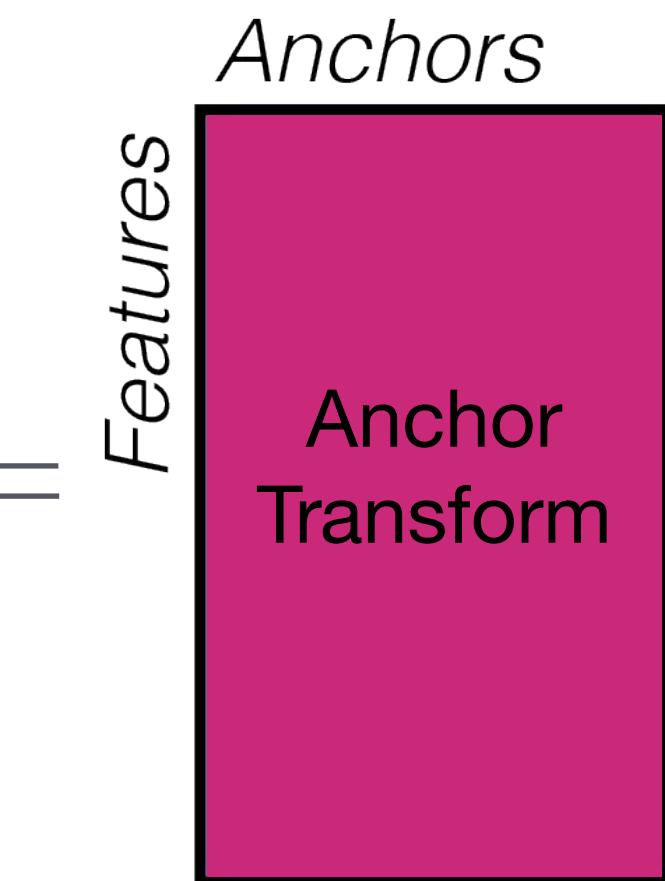
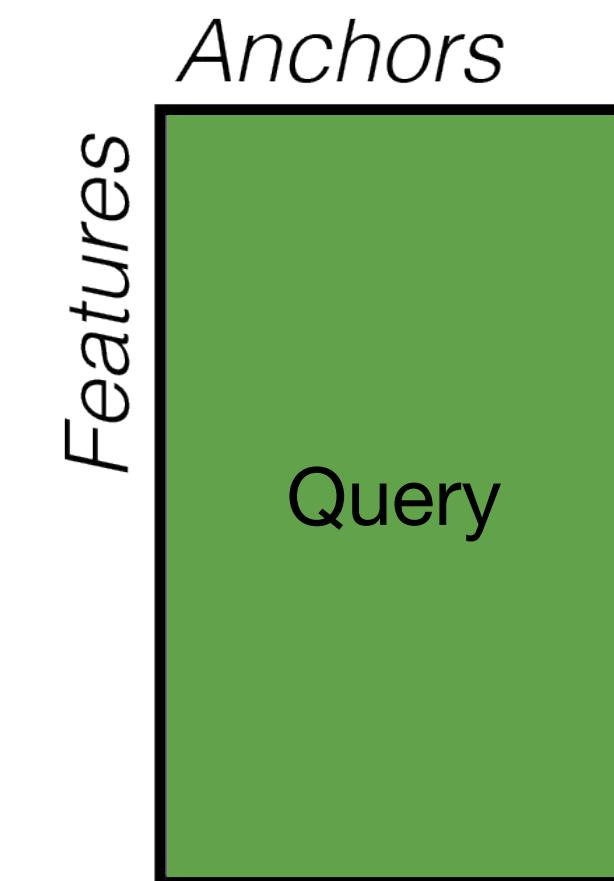
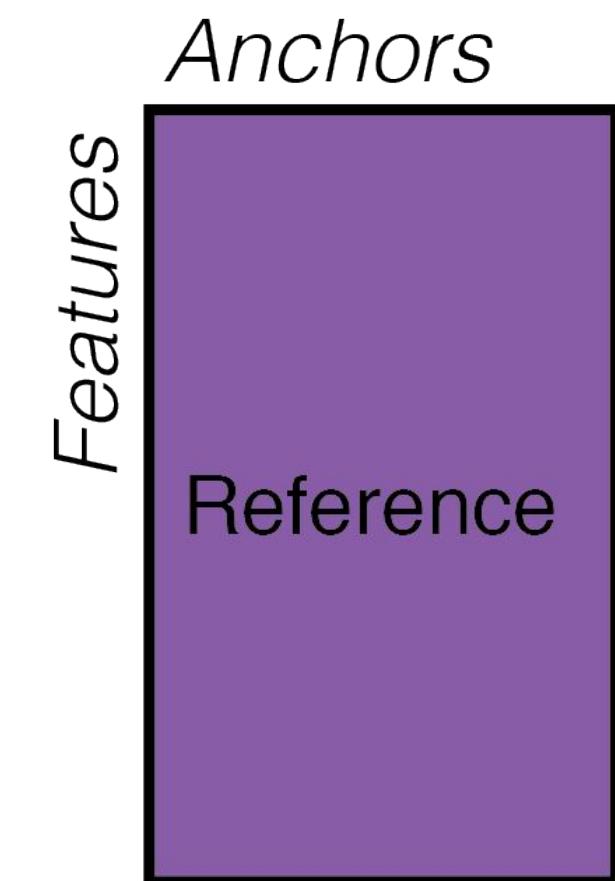


Identifying and filtering incorrect anchors



Incorrect anchors are **infrequent** and
inconsistent

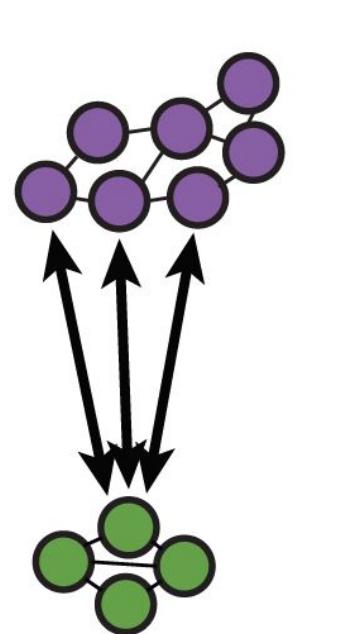
Computing the **integration transformation**



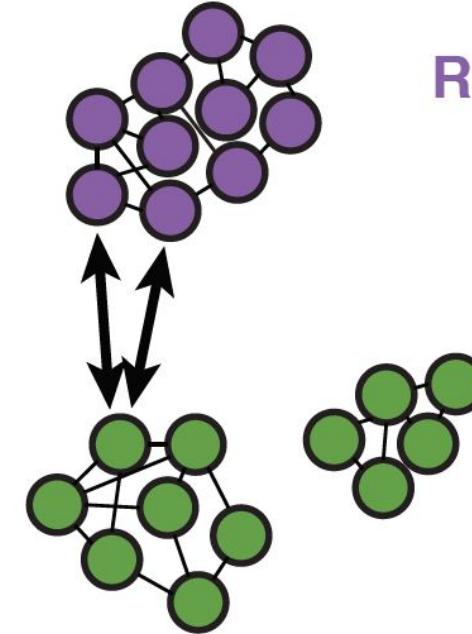
Weighting anchors to reinforce consistency

Confidence score

High-scoring correspondence
Anchors are consistent with local neighborhoods



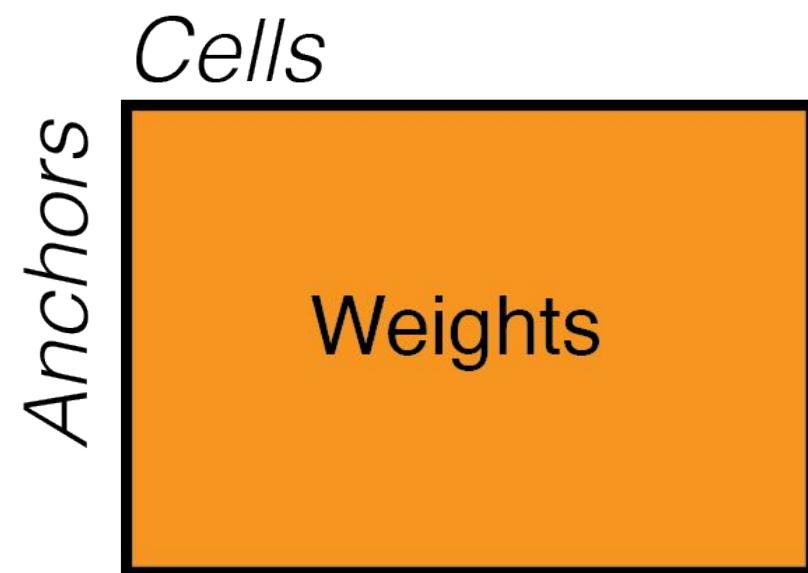
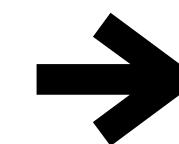
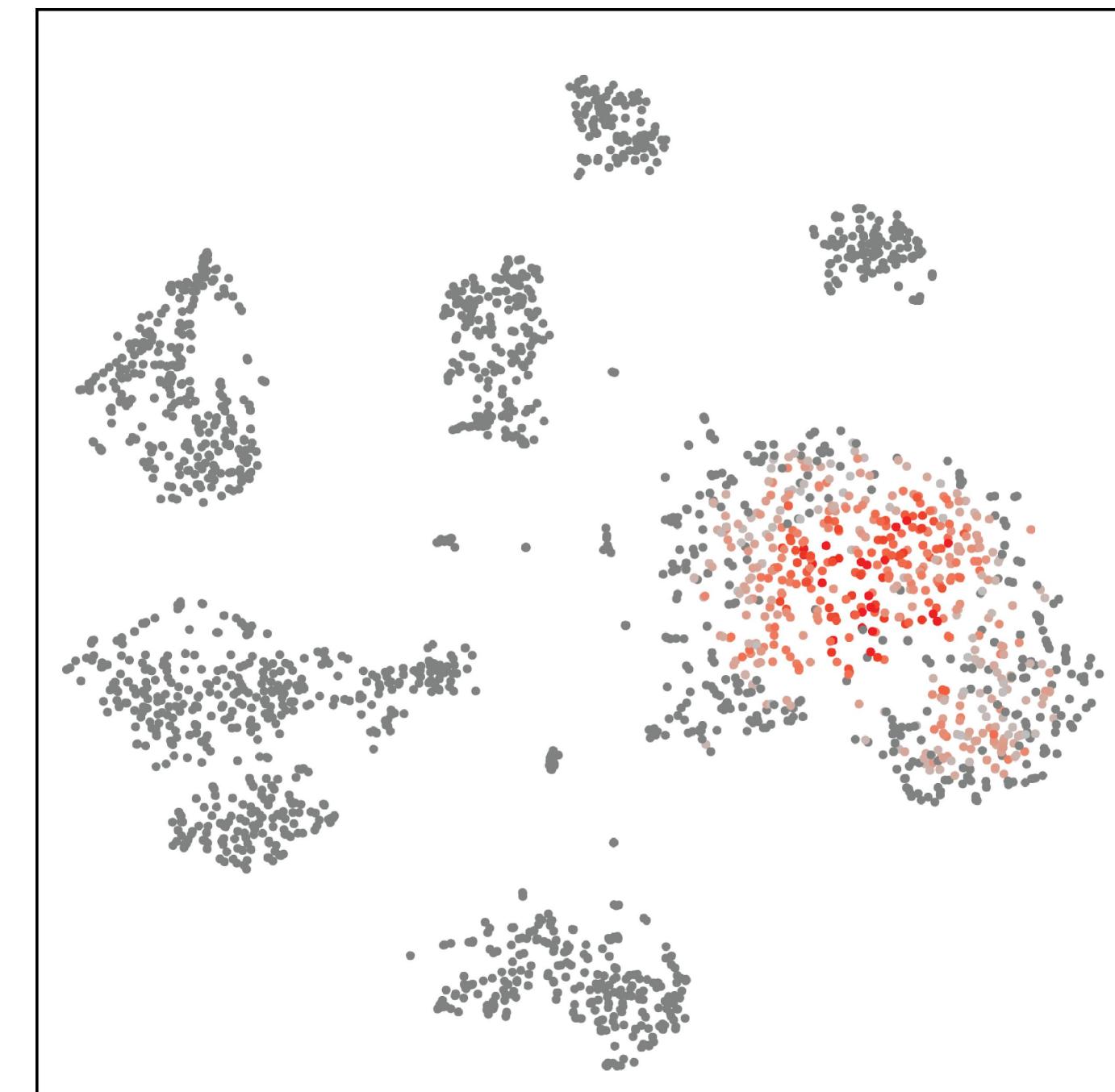
Low-scoring correspondence
Anchors are inconsistent with local neighborhoods



Reference

Query

Cell similarity score

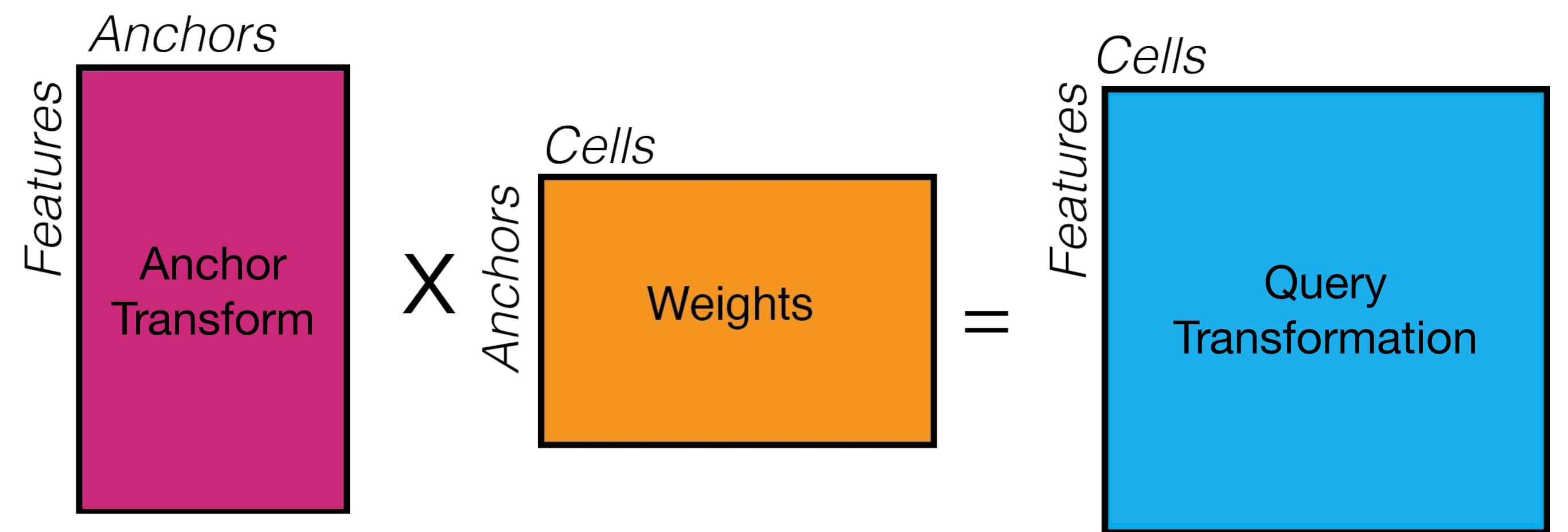


Weight the contribution of each anchor to each cell

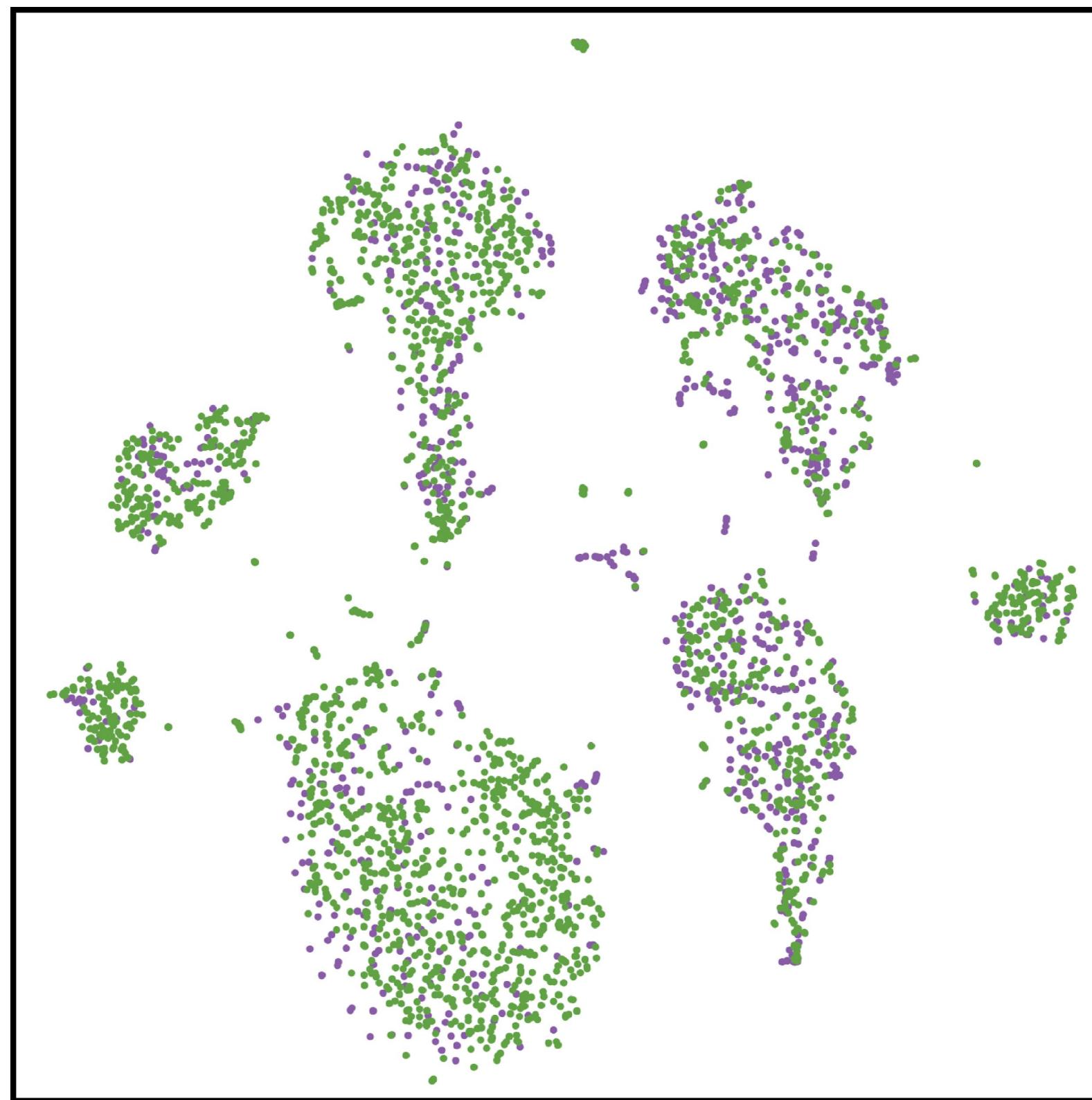
How much do we trust the accuracy of each anchor?

For each cell, which anchor relationships are most relevant?

Integrating datasets



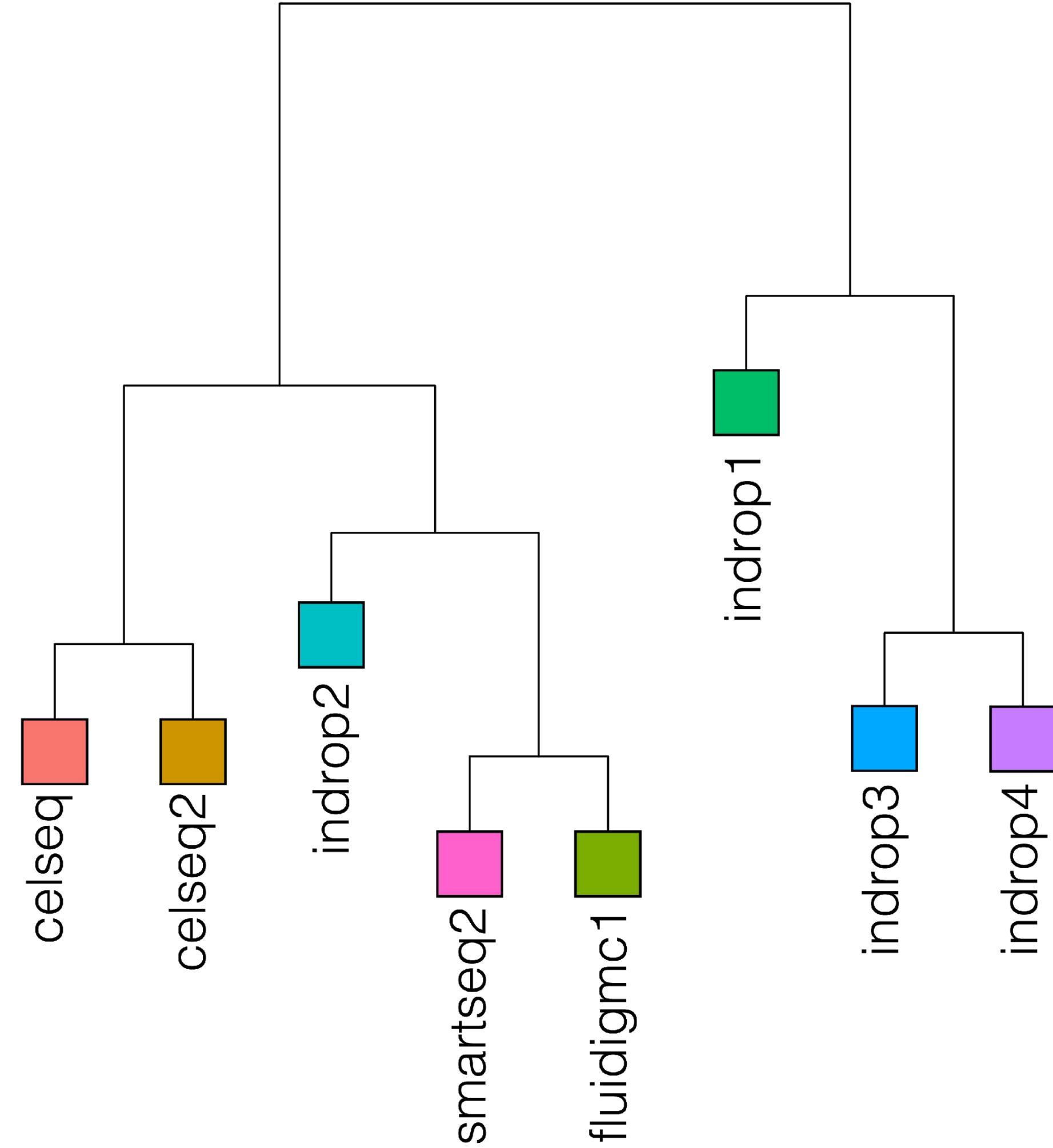
Integrated reference + query



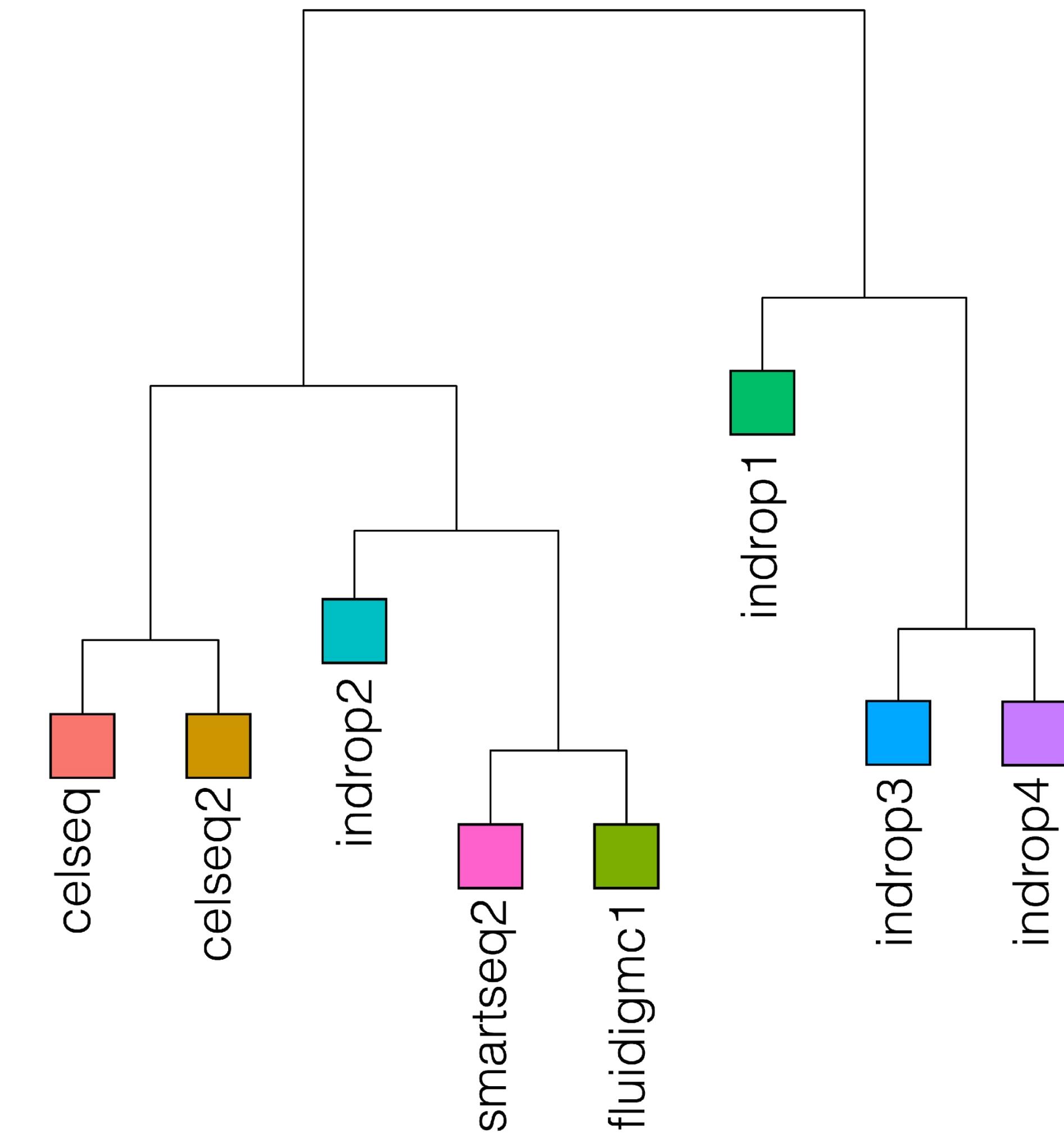
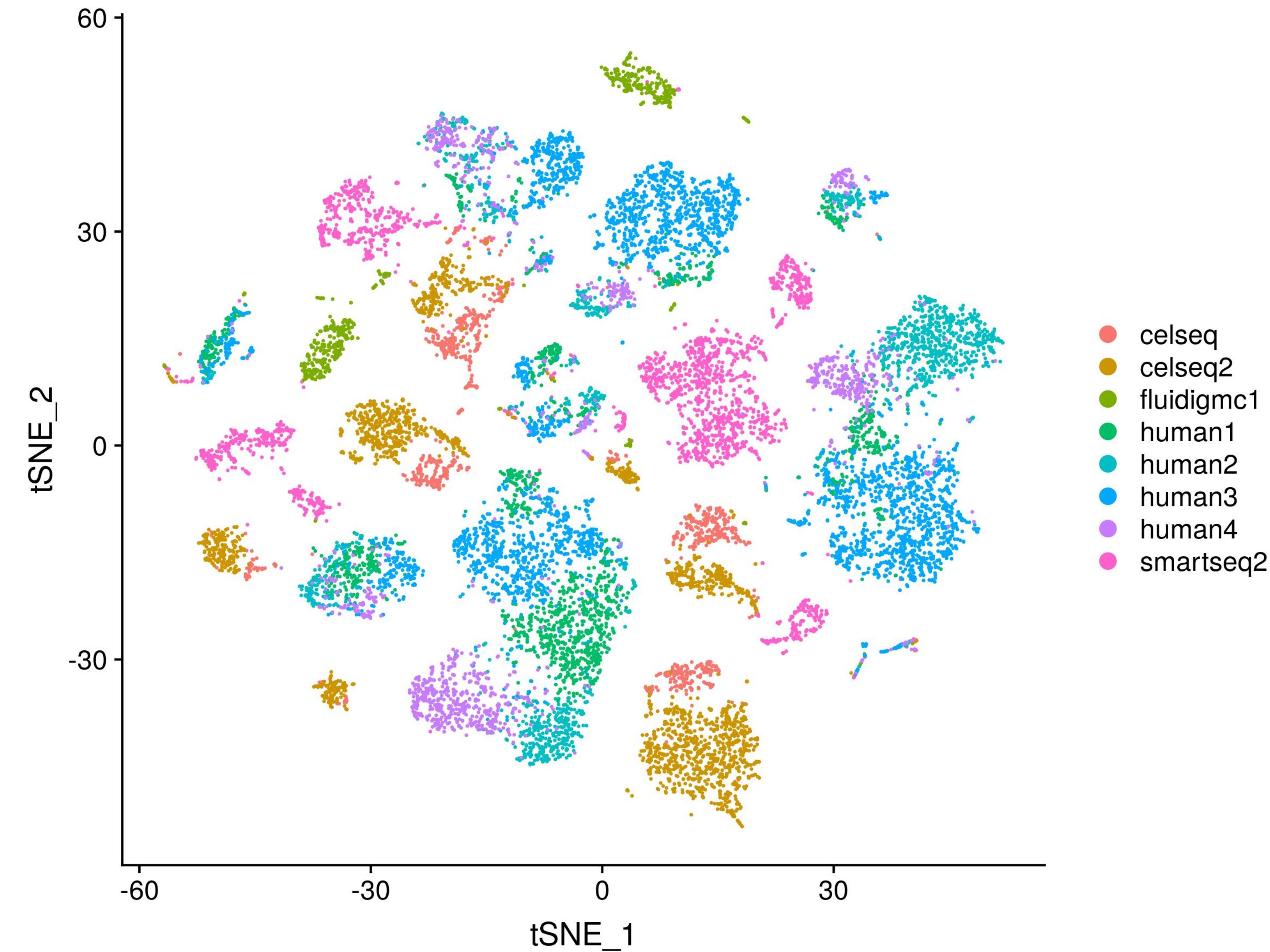
Extension to multi-dataset integration

Build a **guide tree** based on dataset similarities to order pairwise alignments

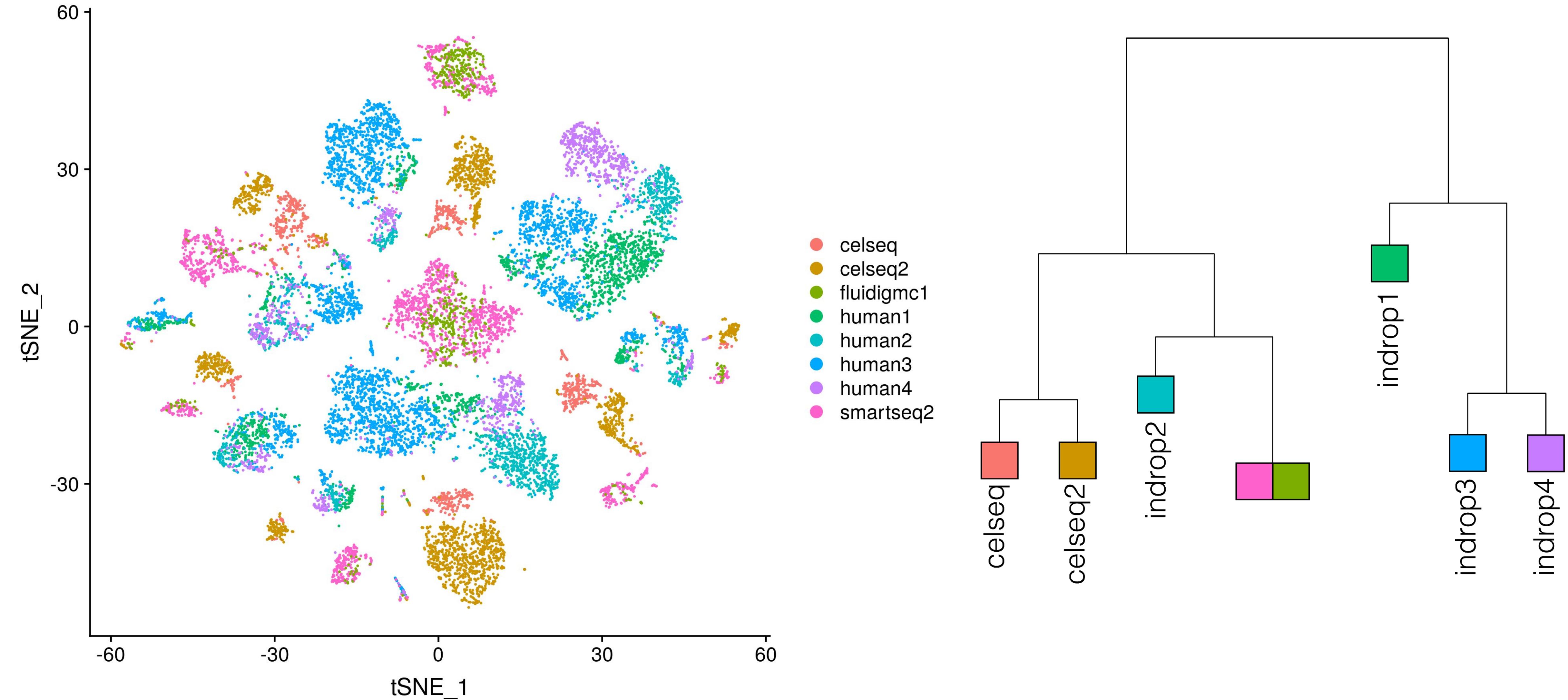
Similar to how many
**multiple sequence
alignment** algorithms work



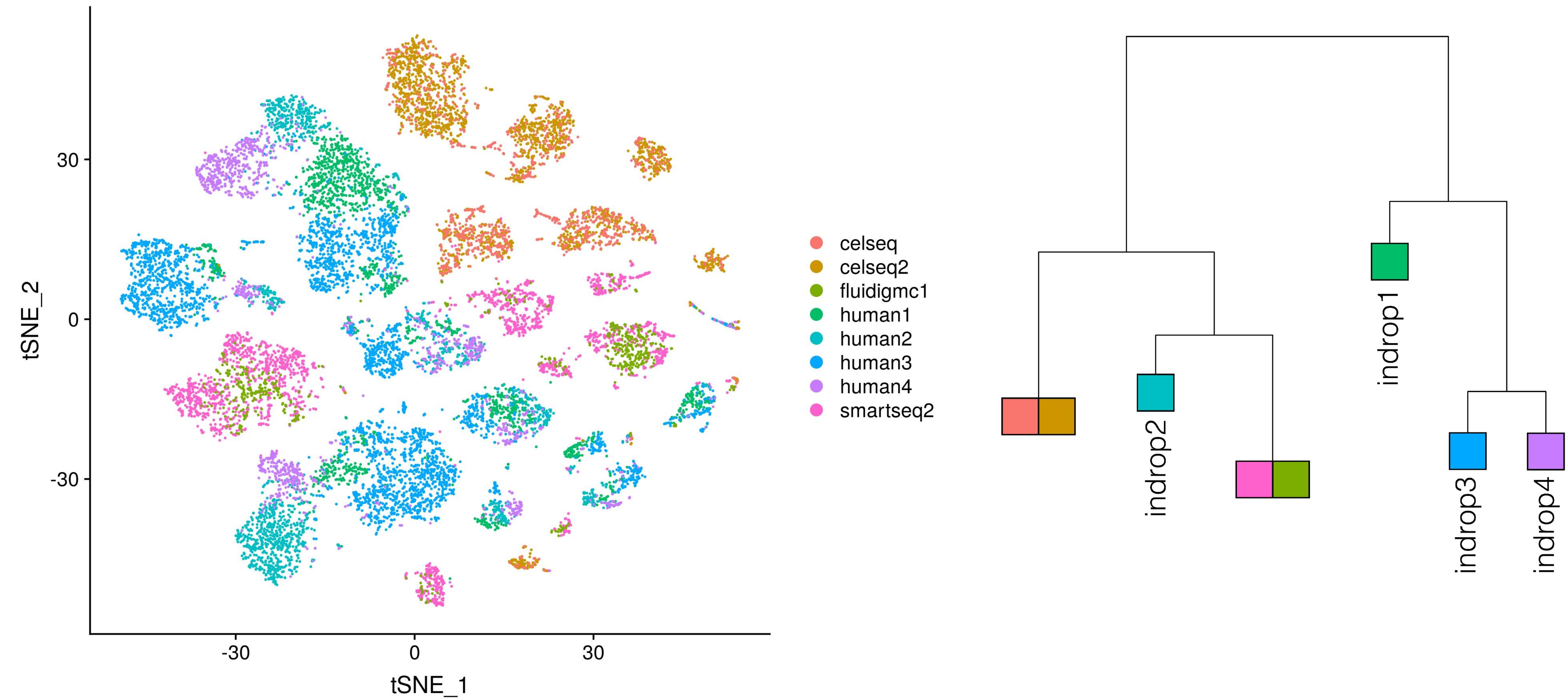
Extension to multi-dataset integration



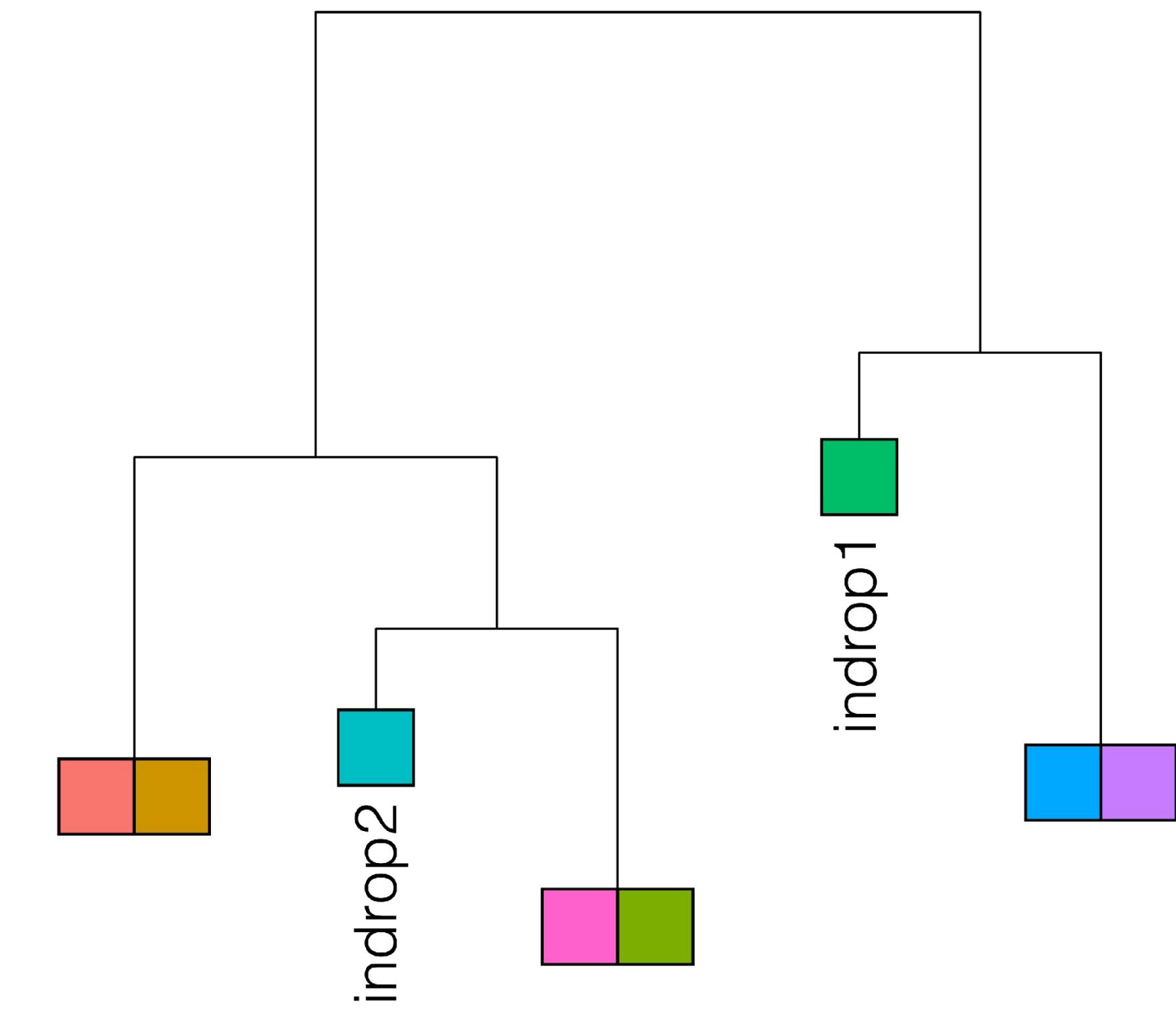
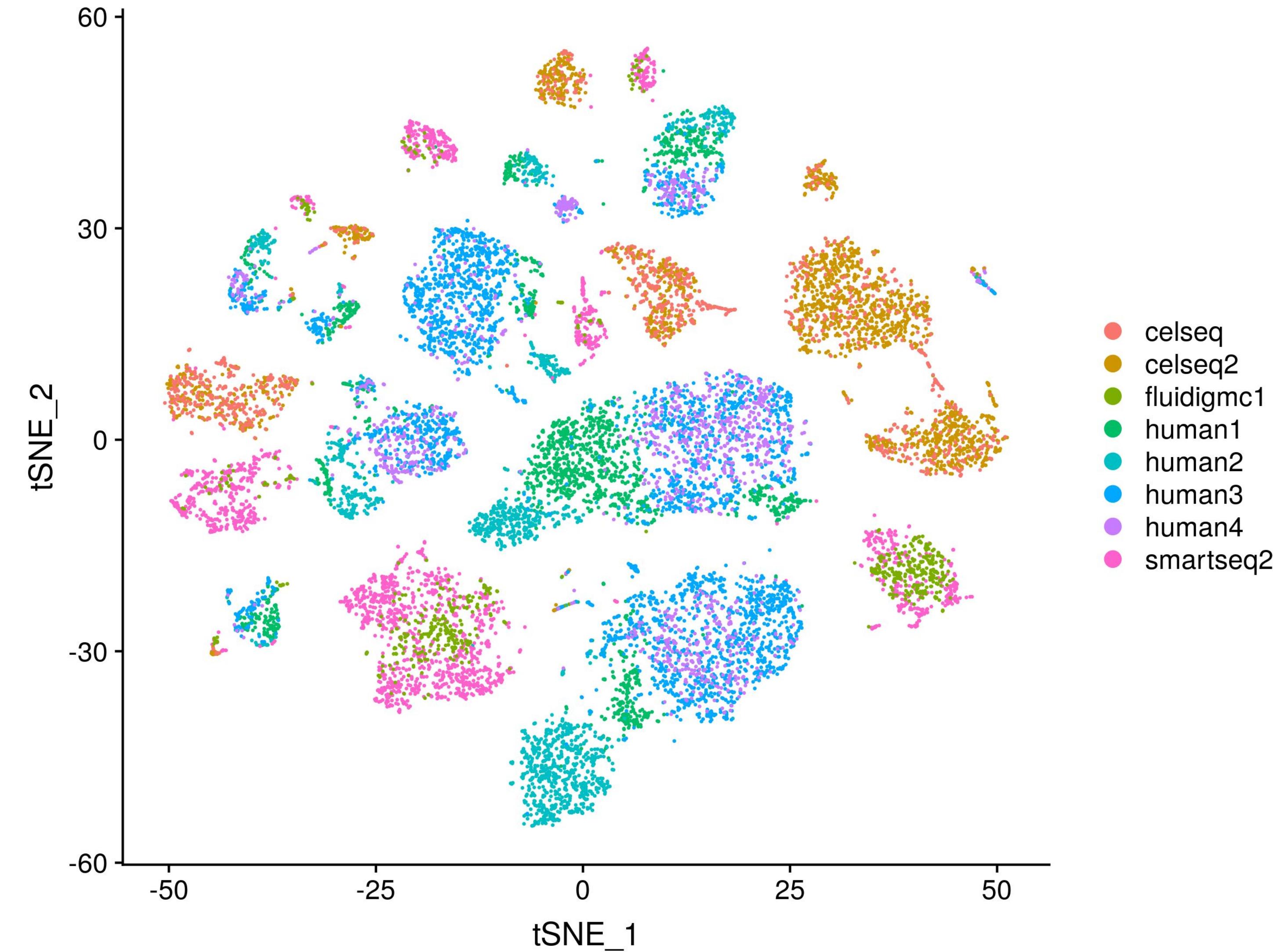
Extension to multi-dataset integration



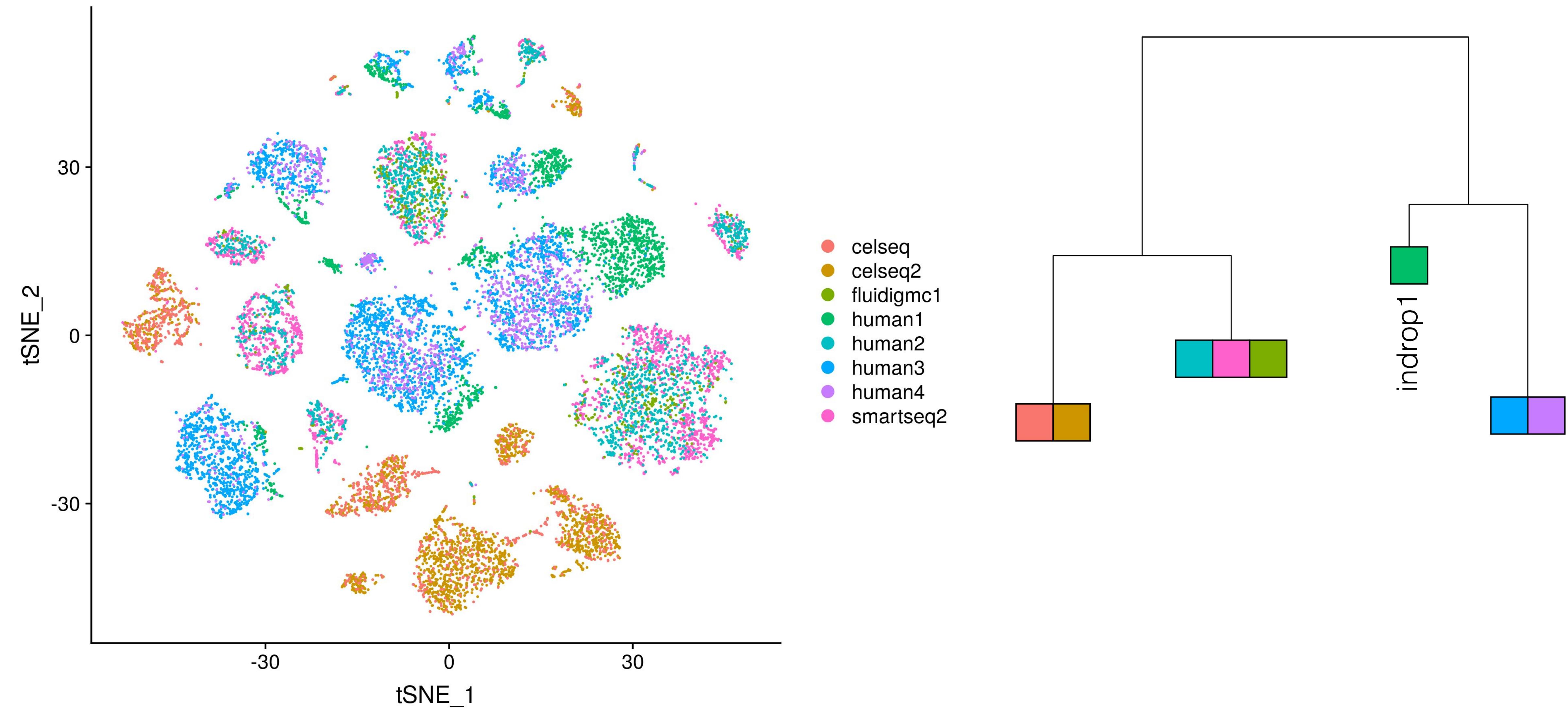
Extension to multi-dataset integration



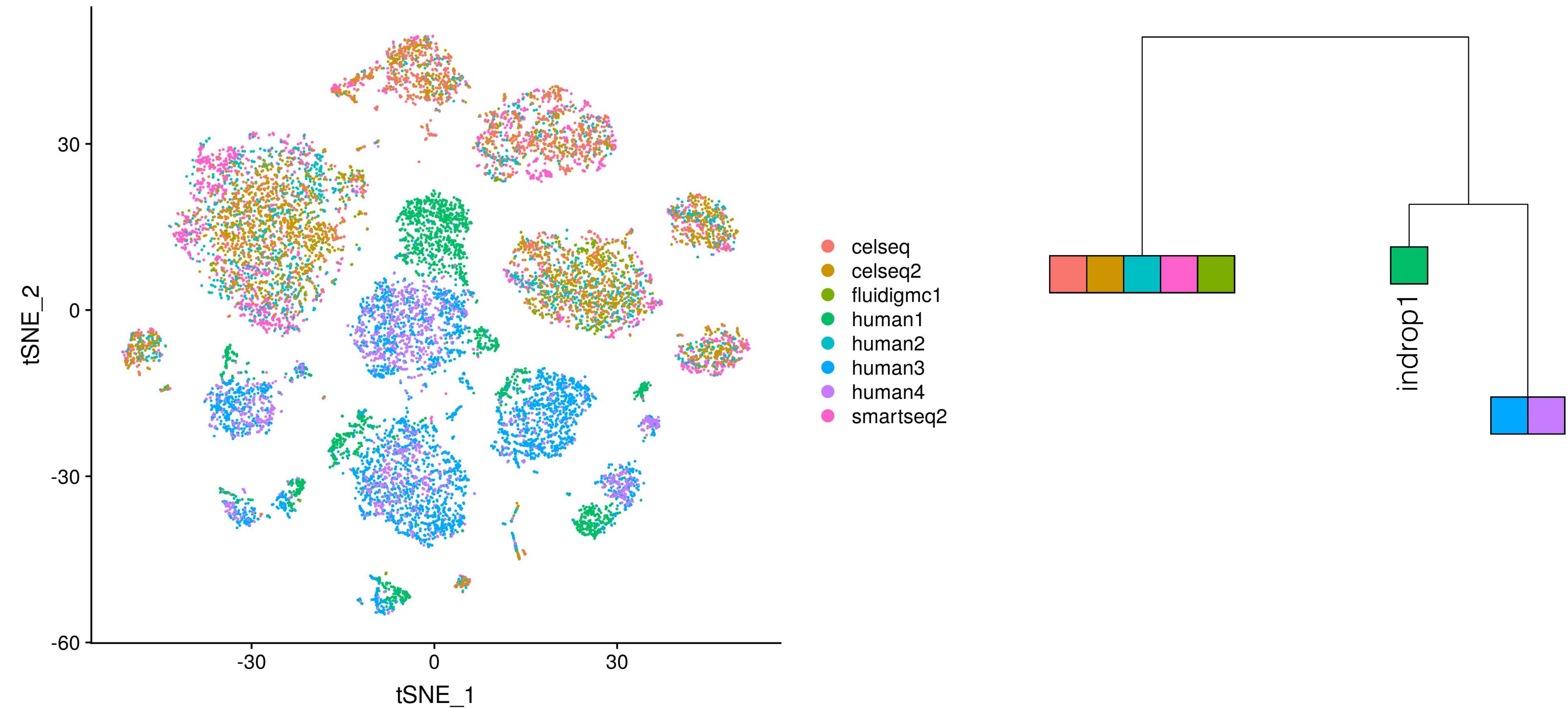
Extension to multi-dataset integration



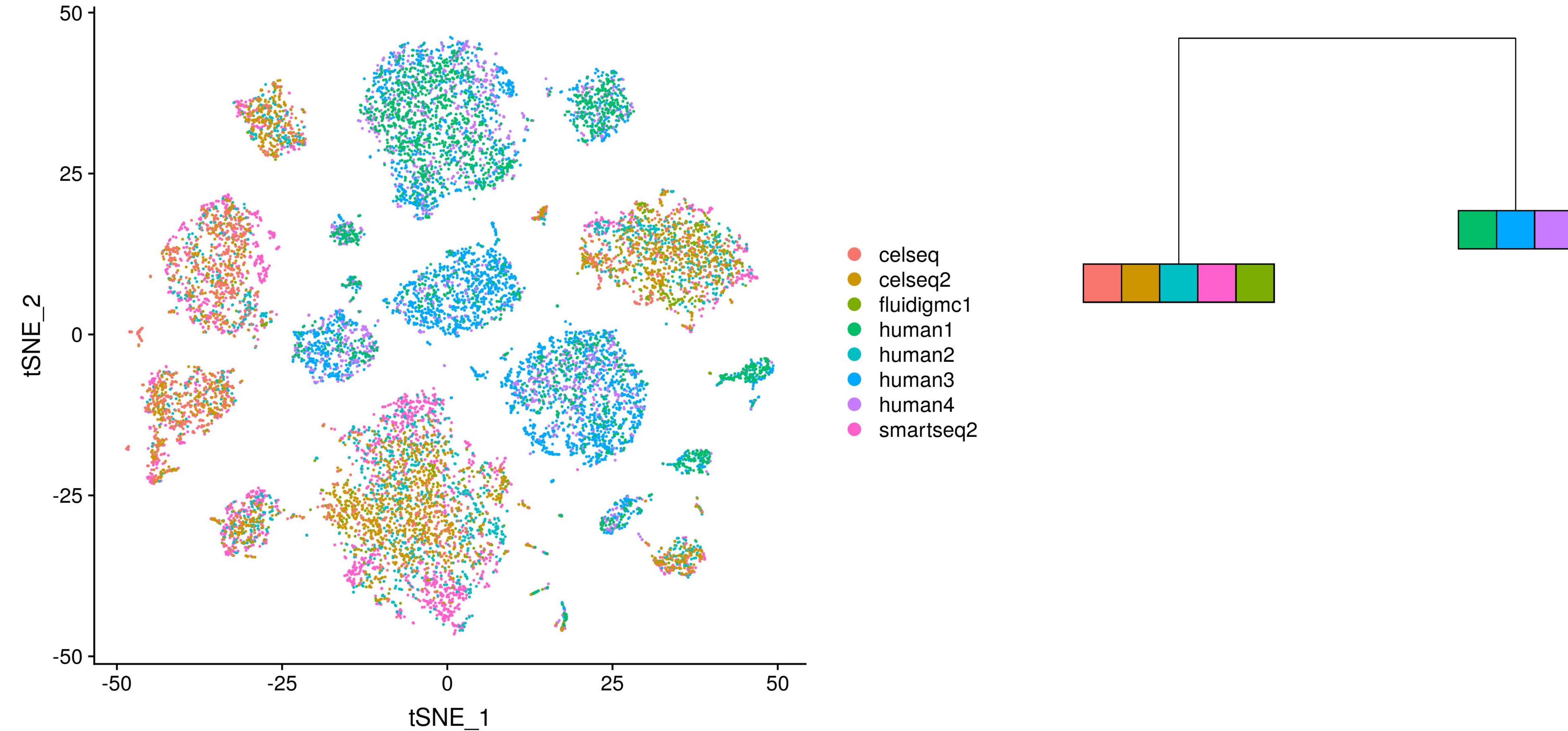
Extension to multi-dataset integration



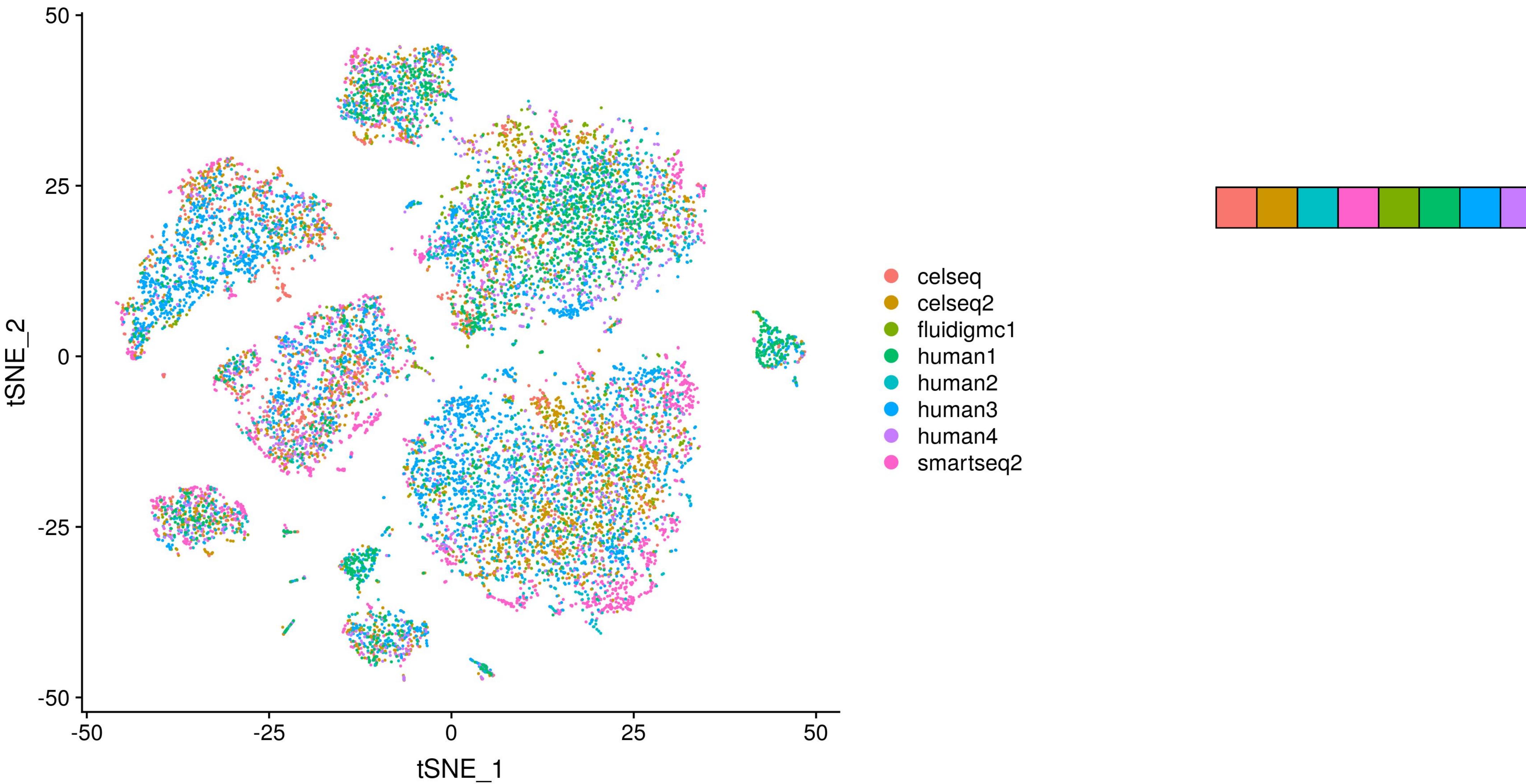
Extension to multi-dataset integration



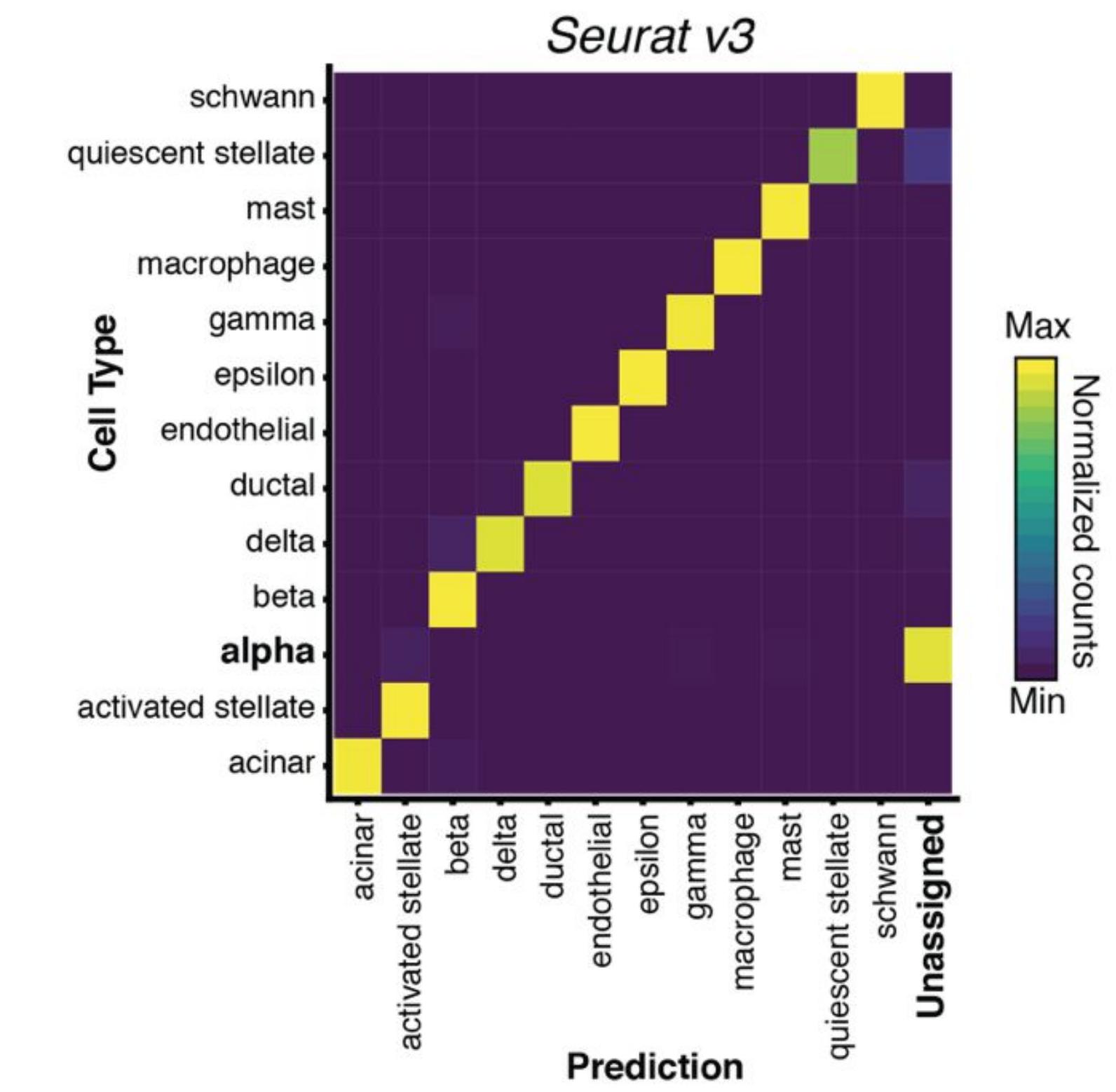
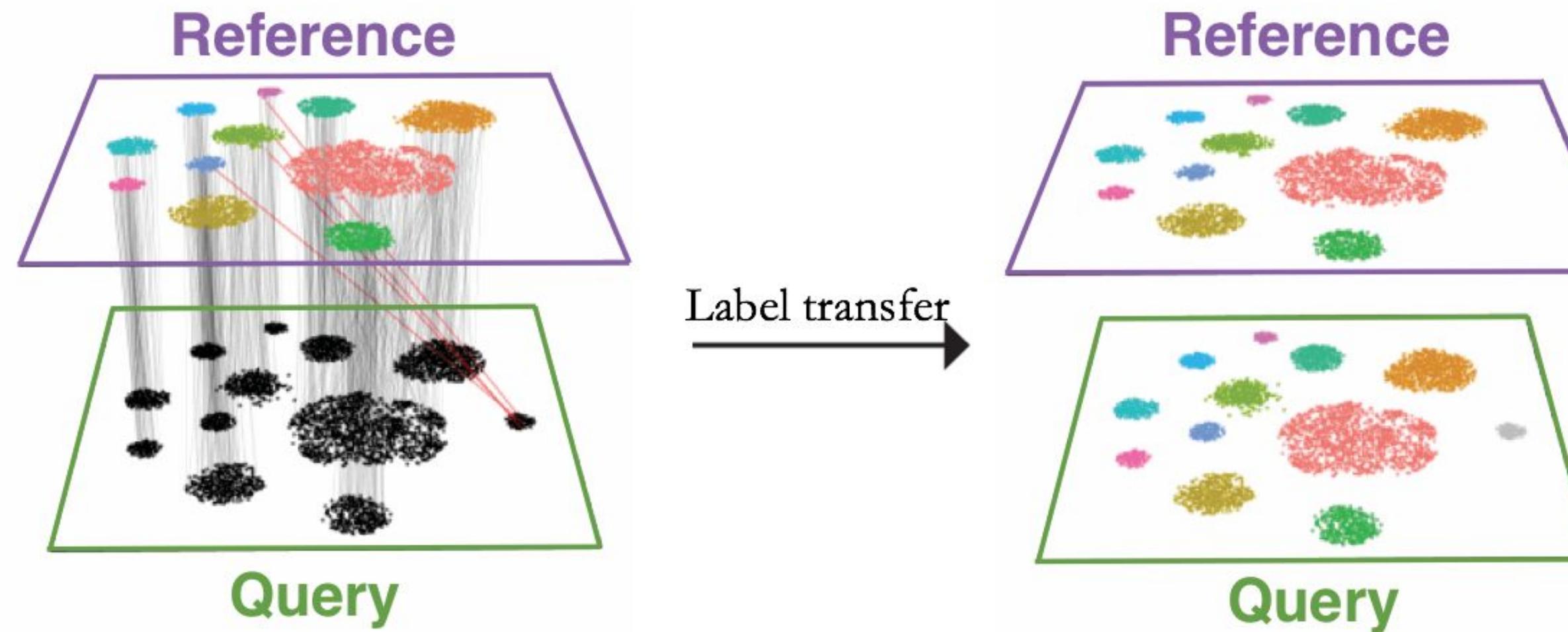
Extension to multi-dataset integration



Extension to multi-dataset integration



Robust identification of anchors enables accurate classifications



Anchors serve as a weighted vote classifier