

Effective Comparison of Single-Cell Embedding Visualizations




Trevor Manz^{1,2}, Fritz Lekschas², Greg Finak², Evan Greene², Nils Gehlenborg¹

¹ Harvard Medical School | ² Ozette Technologies

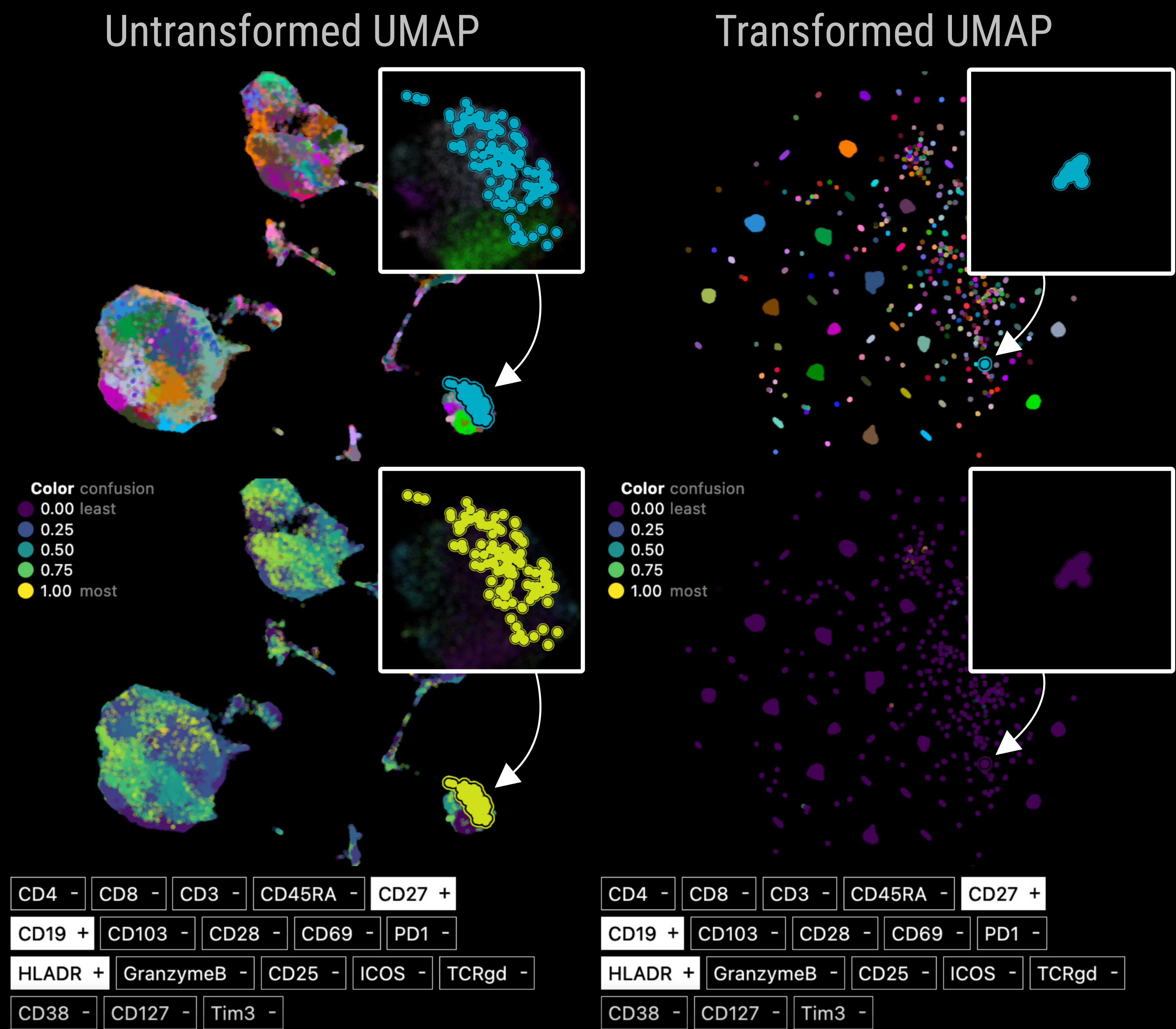
TRY IT YOURSELF:
`pip install cev`

AND CHECK OUT THE EXAMPLES AT:



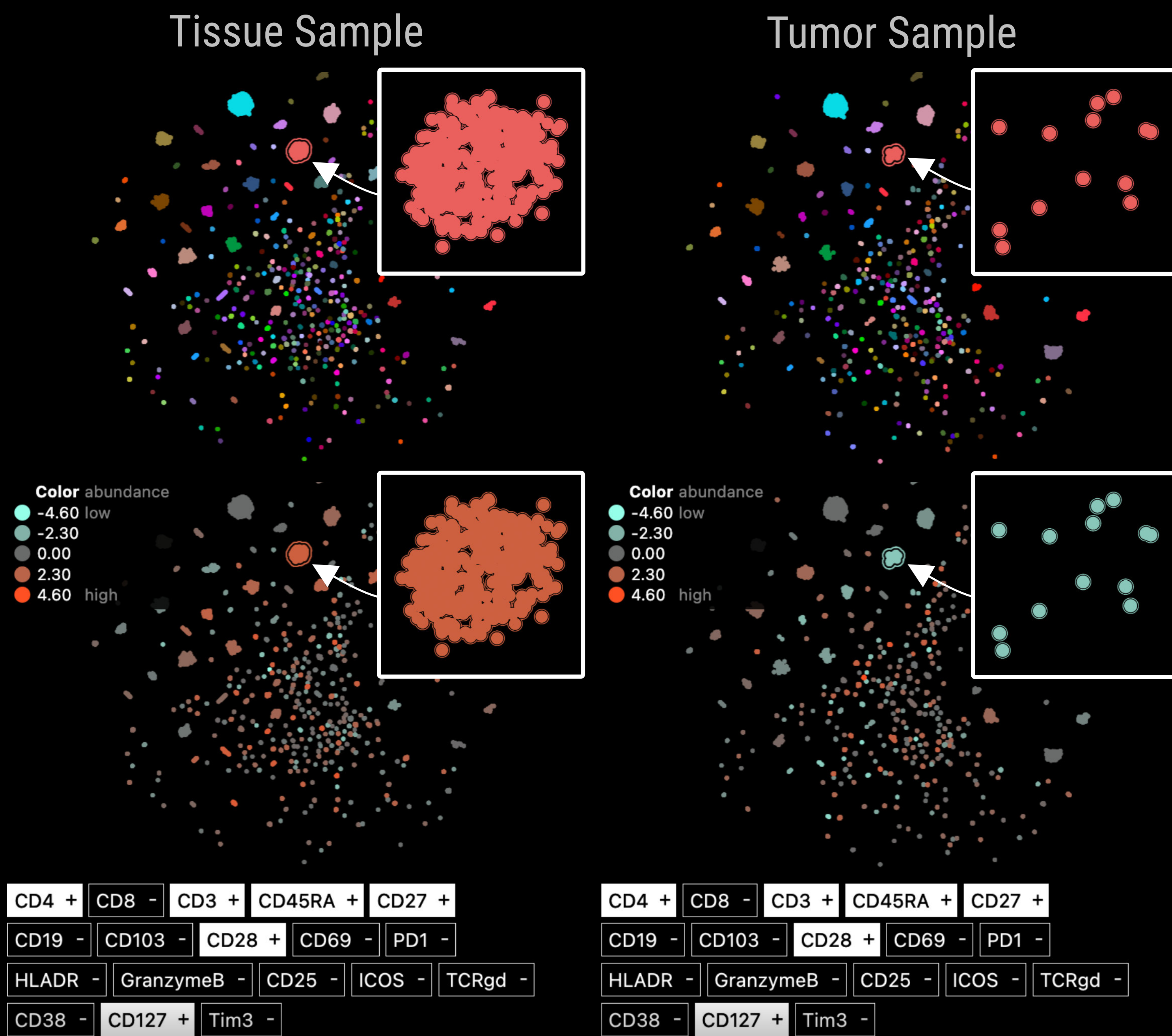


Confusion Between Embedding Methods

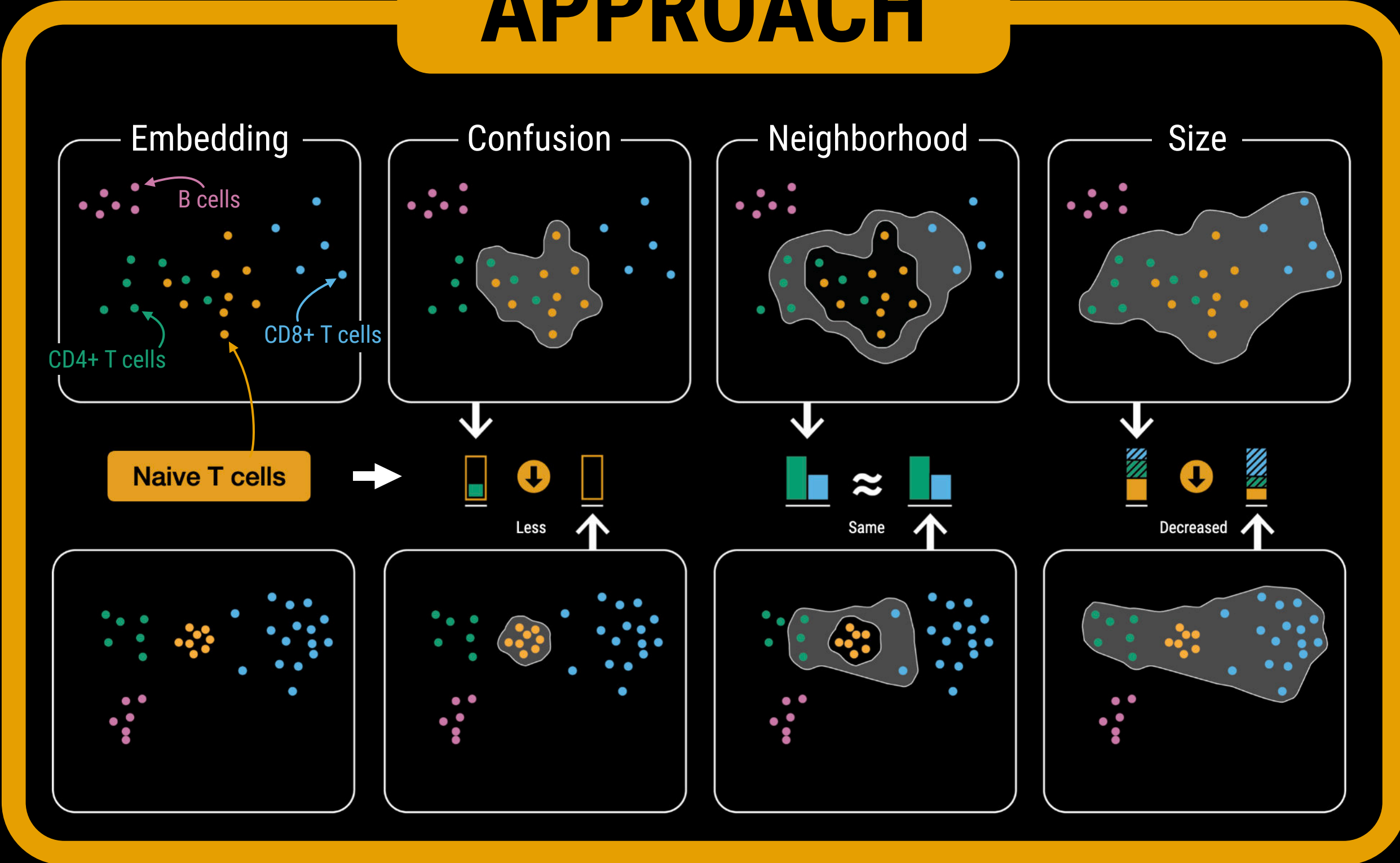


Data from Mair, F. et al., 2022. *Nature*.

Abundance Differences Between Samples



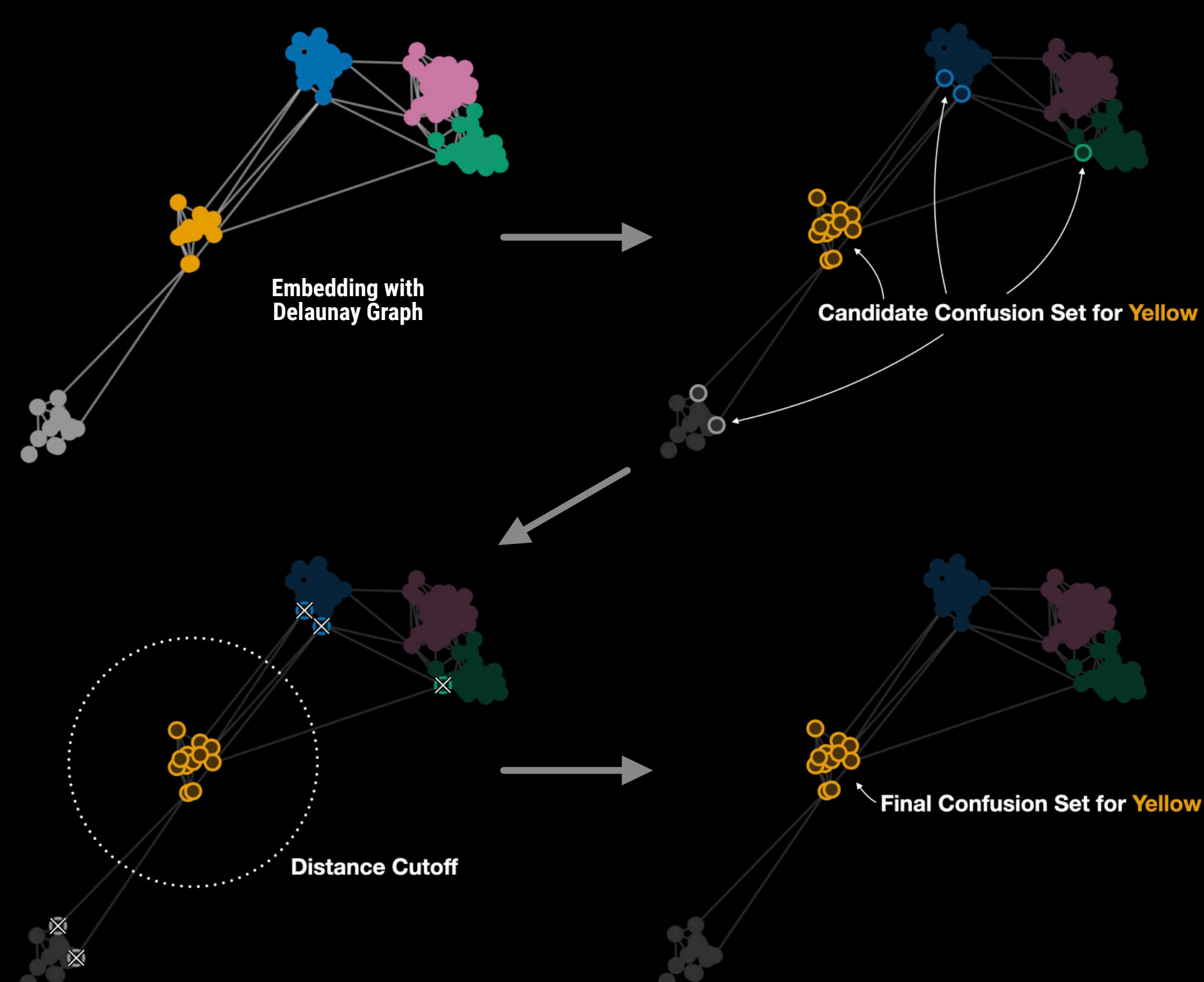
APPROACH



CHALLENGES? Comparing multiple embeddings is a prevalent analysis task; however, existing pairwise comparison methods are constrained to embeddings with shared point correspondences, which do not exist between experiments. Existing metrics also struggle to reveal meaningful similarities and differences when comparing two embeddings of the same dataset.

CONCLUSIONS! We employ set-based similarity metrics on neighborhood graphs from single-cell embeddings, determining the similarity of shared labels at multiple levels between datasets. These metrics guide the visual exploration of such embeddings, enabling quick discovery of interesting phenotypes.

Label Confusion Set



METHODOLOGY

- Create Delaunay graph of the embedded points
- For each label: conduct *breadth-first search* for every point with that label

Construct Confusion Set

- Points within one hop account to label confusion set
- Remove points that are *k* standard deviations away

Construct Neighborhood Set

- Points with 1+ hop and not in the confusion set account for neighborhood set
- Scale neighborhood strength of each label by:
 - Average number of connections between all labels
 - Average distances of connections between all labels

Label Neighborhood Set

