

Data Transformations for Effective Visualization of Single-Cell Embeddings

Evan Greene^{*1}, Greg Finak¹, Fritz Lekschas^{*1}, Malisa Smith^{*1}, Leonard A. D’Amico², Nina Bhardwaj³, Candice D. Church⁴, Chihiro Morishima⁴, Nirasha Ramchurren², Janis M. Taube⁵, Paul T. Nghiem⁴, Martin A. Cheever², Steven P. Fling², and Raphael Gottardo⁶

CONTACT: We’re at ISMB in person. Come say hi! 🙋



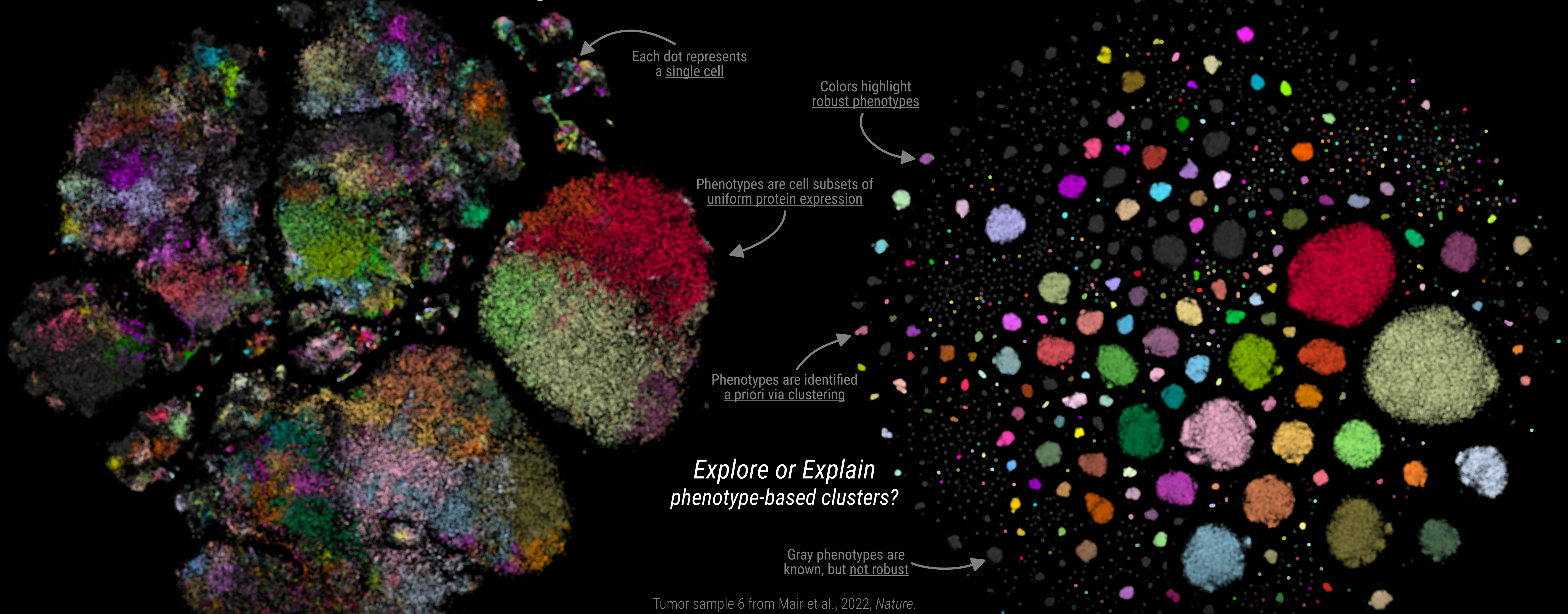
RELATED PAPER:

Evan Greene et al., 2021. New interpretable machine-learning method for single-cell data reveals correlates of clinical response to cancer immunotherapy. *Patterns*.



Standard Embedding

Annotation Embedding




WHAT? Single-cell embeddings are lower-dimensional (often 2D) representations of high-dimensional cellular data that try to capture the similarity of cells in terms of their cellular features. Such features can come from the proteome (e.g., surface proteins), genome (e.g., genes), or epigenome (e.g., CREs, TFBS, or chromatin accessibility peaks). The hope is that cells with similar features are located close to each other.

WHY? The goal of visualizing such embeddings is often to obtain a complete overview of all cells and how they relate to each other to generate hypothesis, annotate clusters, or compare cell types.

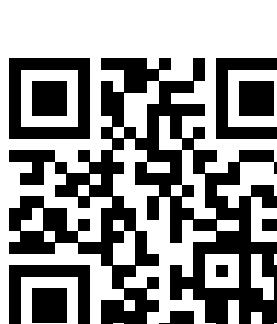
METHODS

Annotate Expressions:

Define expression levels
E.g., positive and negative




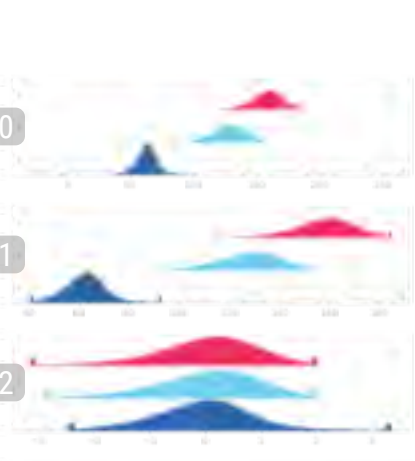
→ Fully interpretable clusters



Code of the clustering method.

Transform Expressions:

1. Remove outlier expressions
Winsorize to 1%–99%
2. Remove marker differences
Normalize to zero mean unit variance
3. Align marker expressions
Translate mean by fixed value



Code of the transformation method.

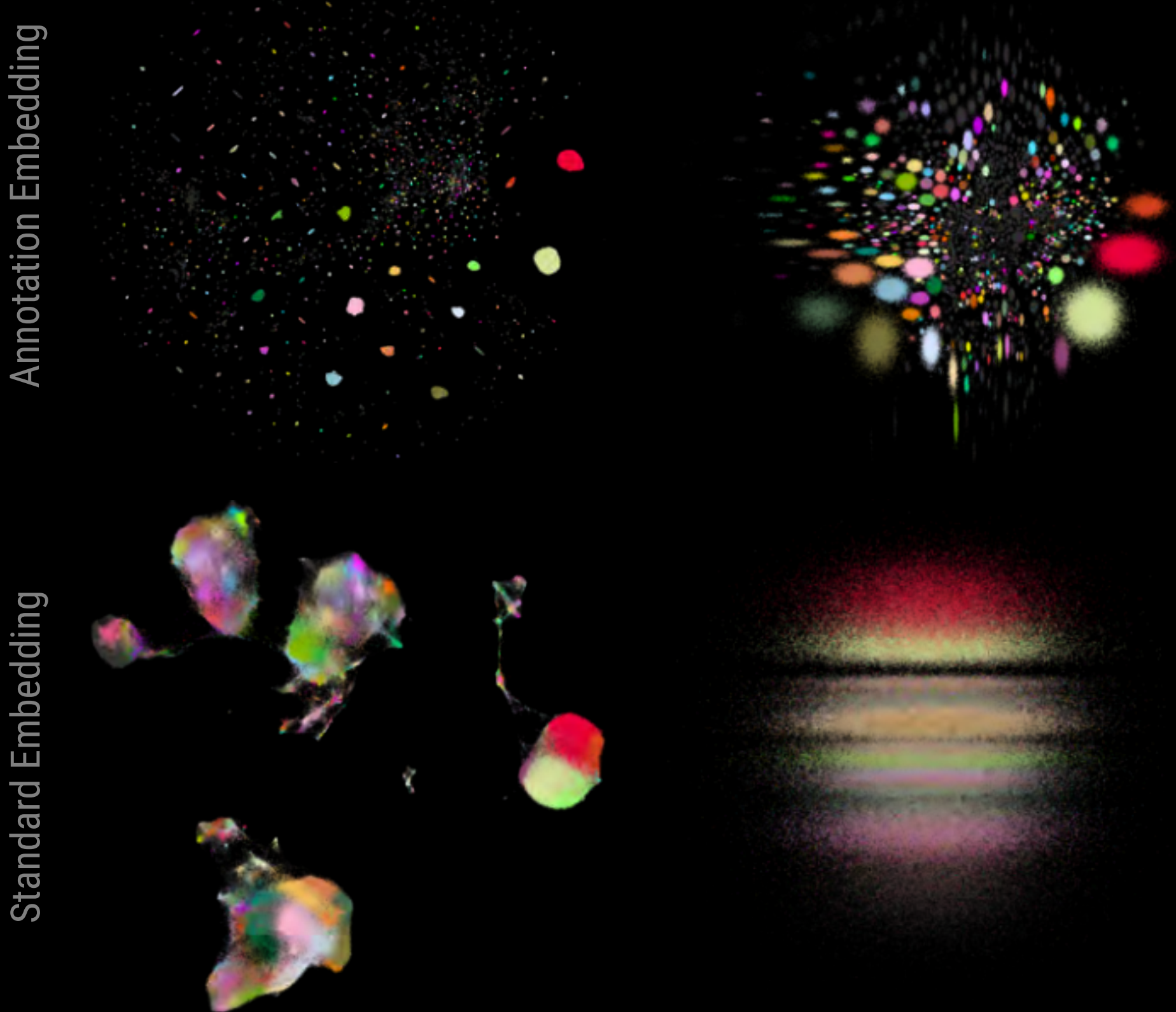
CHALLENGES? While standard embeddings offer exciting opportunity for exploration, they might not be able to resolve complex cellular phenotypes or adjust for batch effects. Our annotation embedding addresses these challenges by transforming the cellular features prior to embedding using expression level annotations.

CONCLUSIONS! "Tune" your data before embedding it. Use a data representation that is closest to your visualization task. Our annotation transformation is useful for explaining clusters by improving cluster coherence and adjusting for batch effects.

Method Independent

UMAP:

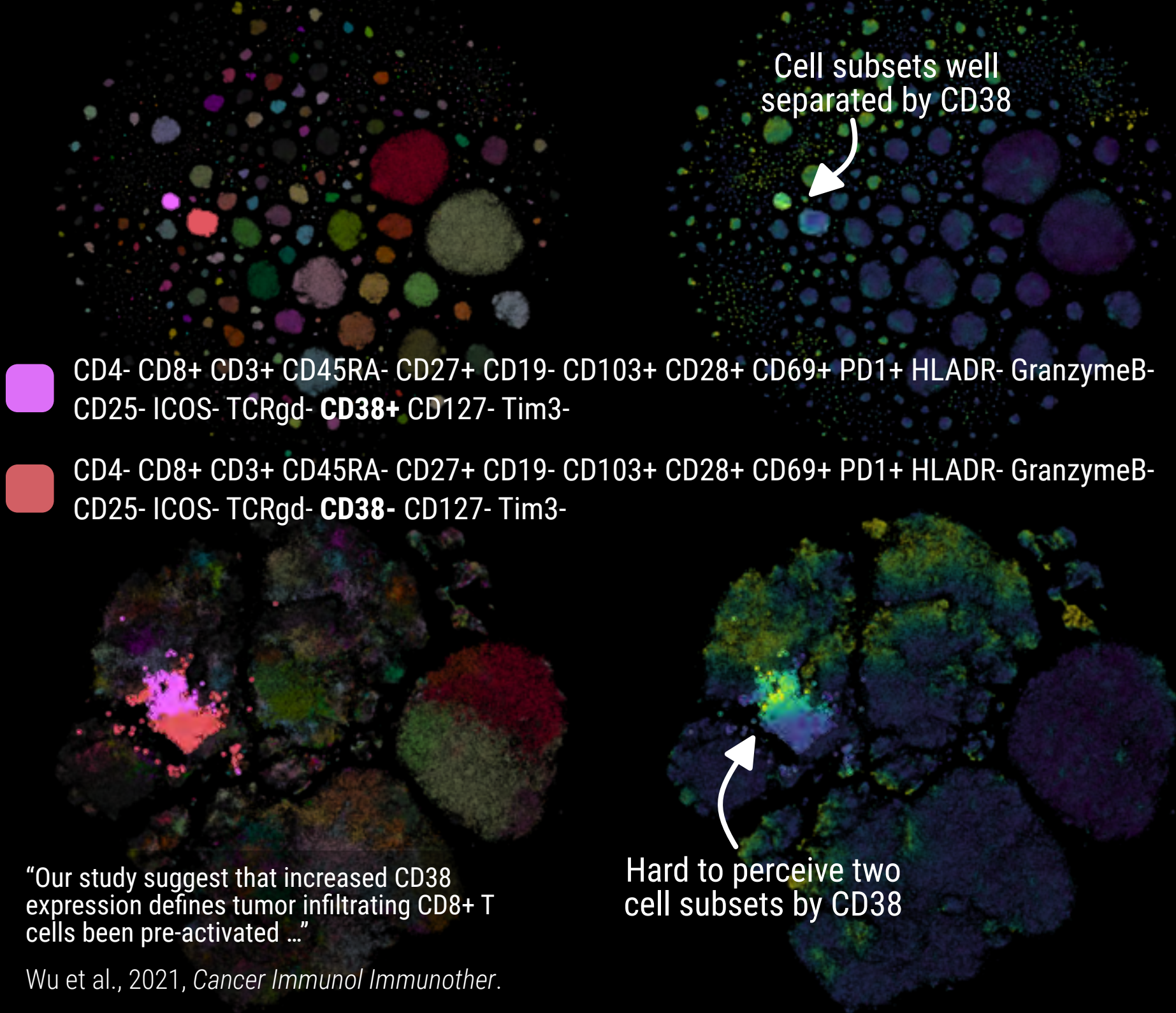
VAE:



Cluster Coherence

Phenotypes:

CD38 Expression:



Batch Effect Adjustment

Joint Embedding:

