

Data Transformations for Effective Visualization of Single-Cell Embeddings

PRESENTER

Fritz Lekschas | Research Scientist at Ozette Technologies

AUTHORS

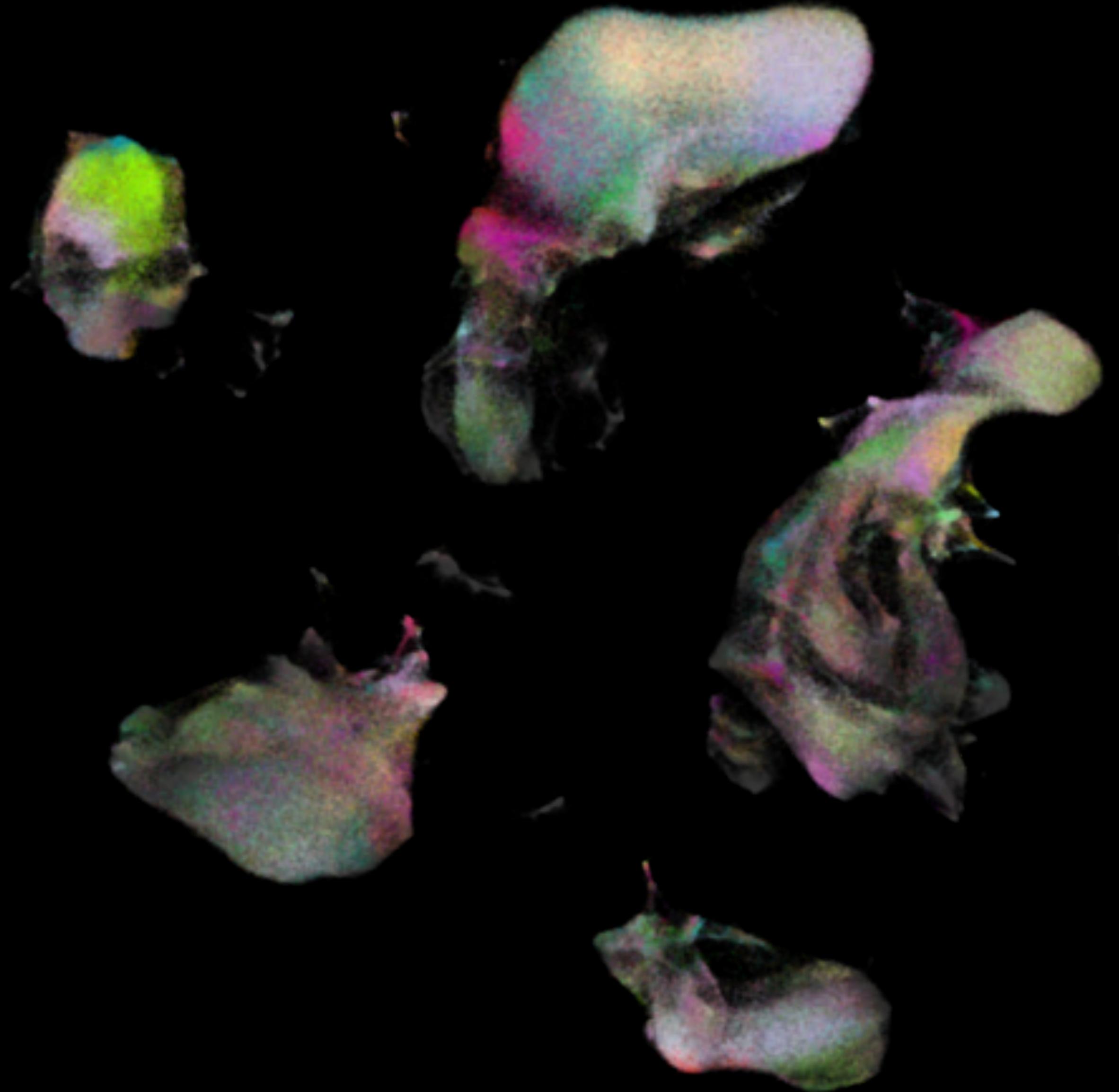
Evan Greene, Greg Finak, Fritz Lekschas, Malisa Smith, Leonard A. D'Amico, Nina Bhardwaj, Candice D. Church, Chihiro Morishima, Nirasha Ramchurren, Janis M. Taube, Paul T. Nghiem, Martin A. Cheever, Steven P. Fling, and Raphael Gottardo



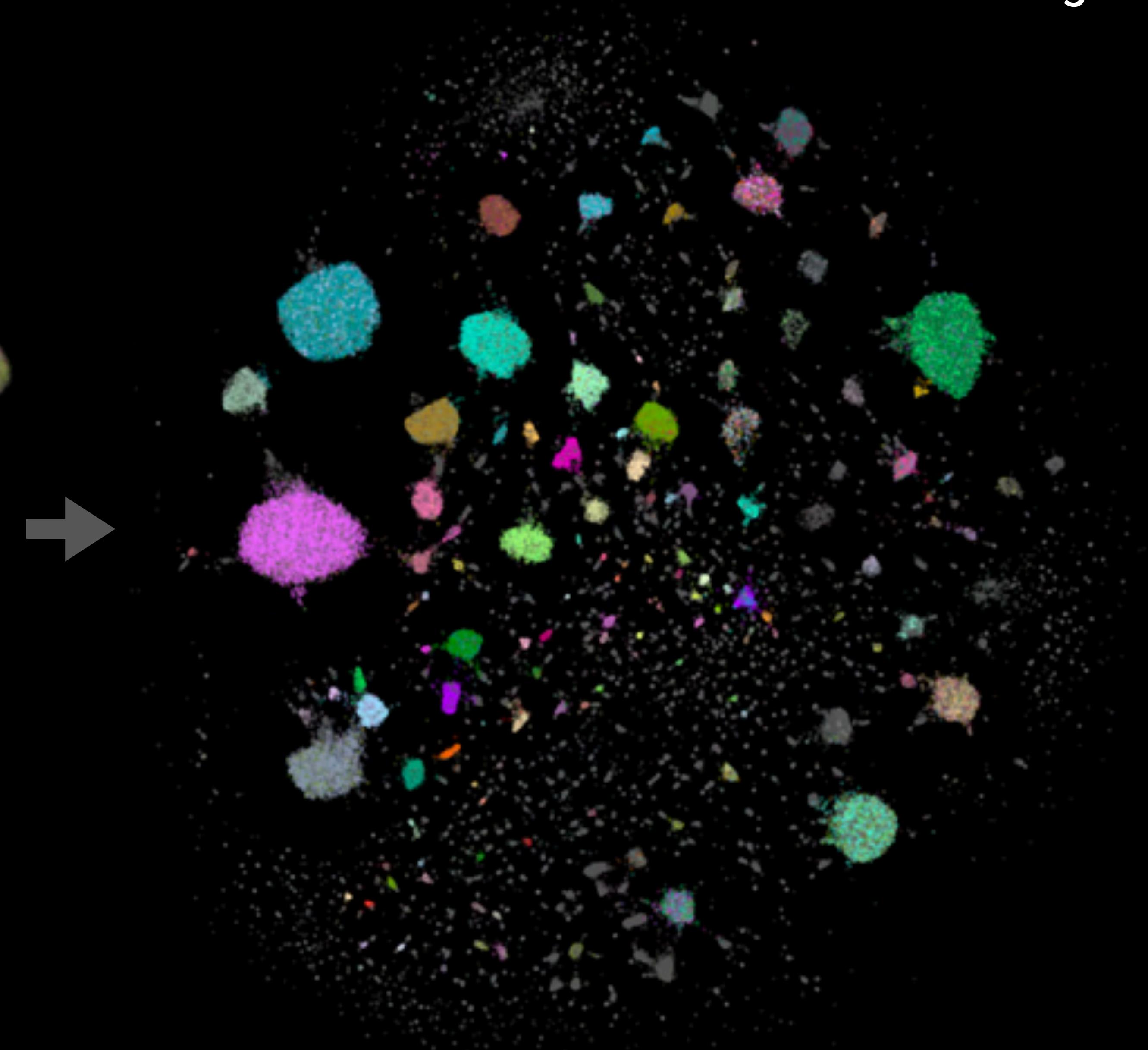
DISCLAIMER

Several authors are affiliated Ozette Technologies.
However, the work had been conducted at Fred Hutch,
has been peer-reviewed, is published, and can be
accessed on the web freely and openly.

Standard Embedding



Annotation-Transformation Embedding



Data from Subrahmanyam et al., 2018, *Journal for ImmunoTherapy of Cancer*.

Single-Cell Embeddings

1. WHAT?

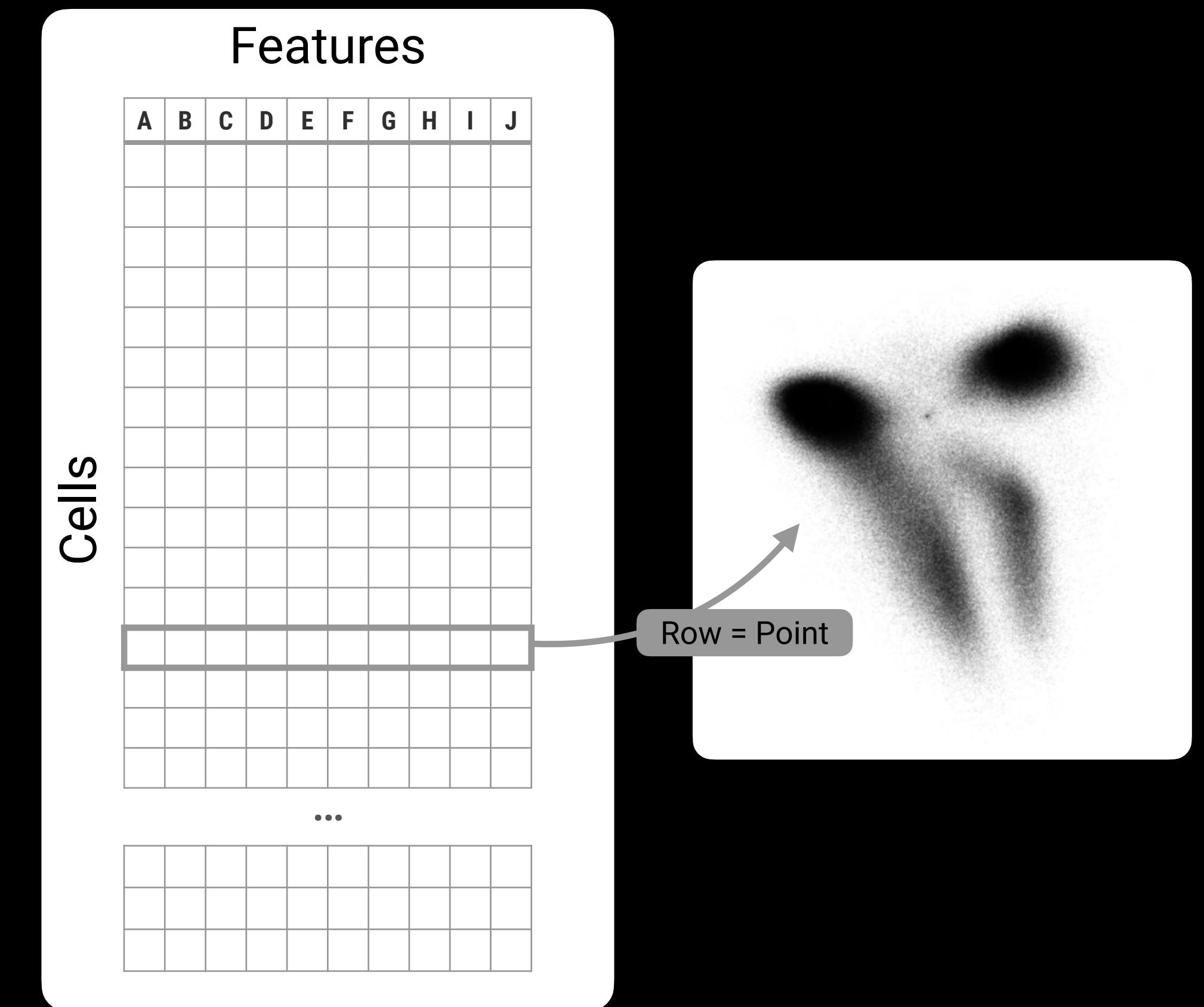
- Low-dimensional representation
- High-dimensional single-cell data
- Proteomics, genomics, epigenomics

2. WHY?

- Gain an overview of all cell populations
- Compare and cluster cells

3. HOW?

- Linear (e.g., PCA)
- Non-Linear (e.g., *t-SNE*, UMAP, or VAE)



Single-Cell Embeddings

1. WHAT?

- Low-dimensional representation
- High-dimensional single-cell data
- Proteomics, genomics, epigenomics

2. WHY?

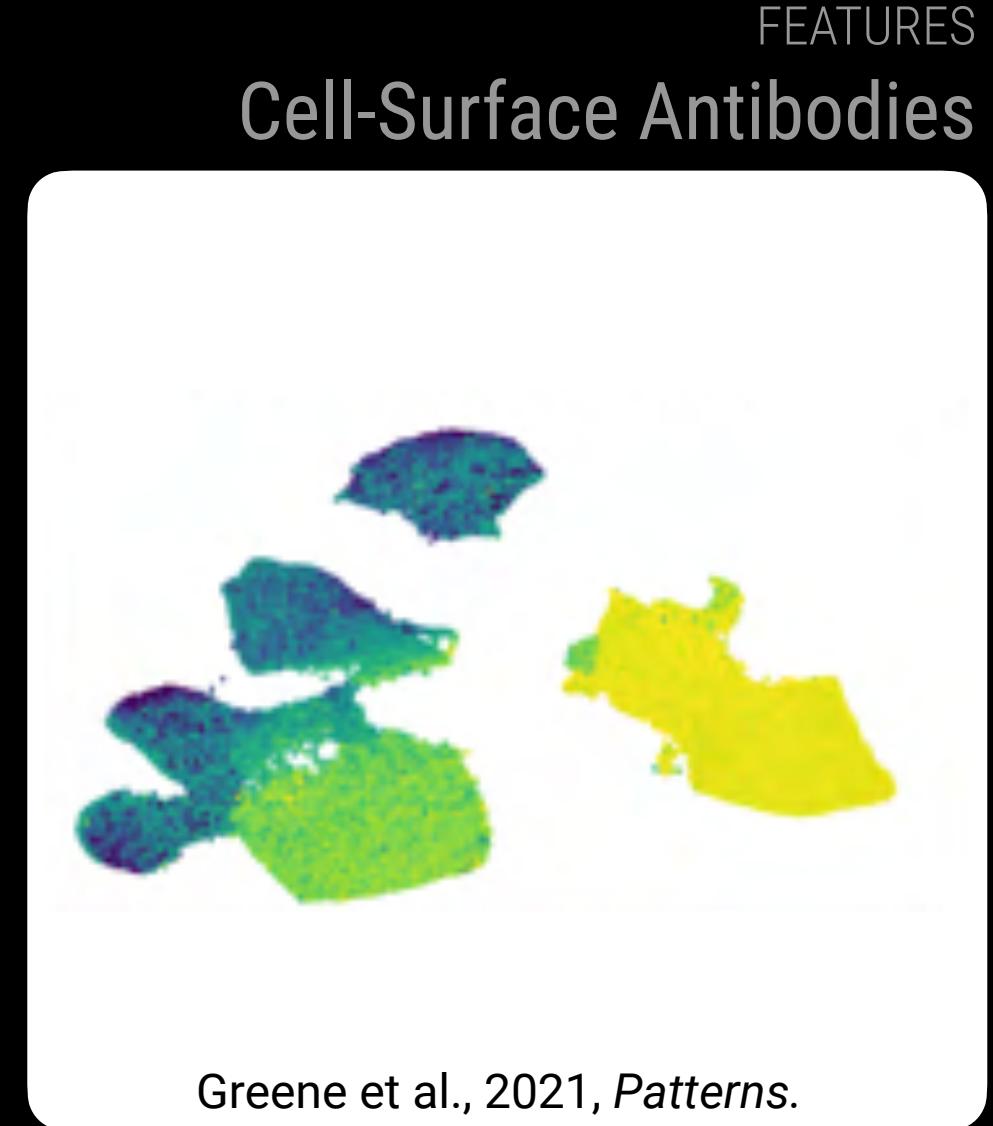
- Gain an overview of all cell populations
- Compare and cluster cells

3. HOW?

- Linear (e.g., PCA)
- Non-Linear (e.g., *t-SNE*, UMAP, or VAE)



Tabula Sapiens Consortium, 2022, *Science*.



Greene et al., 2021, *Patterns*.



Granja et al., 2020, *Nature Biotechnology*.

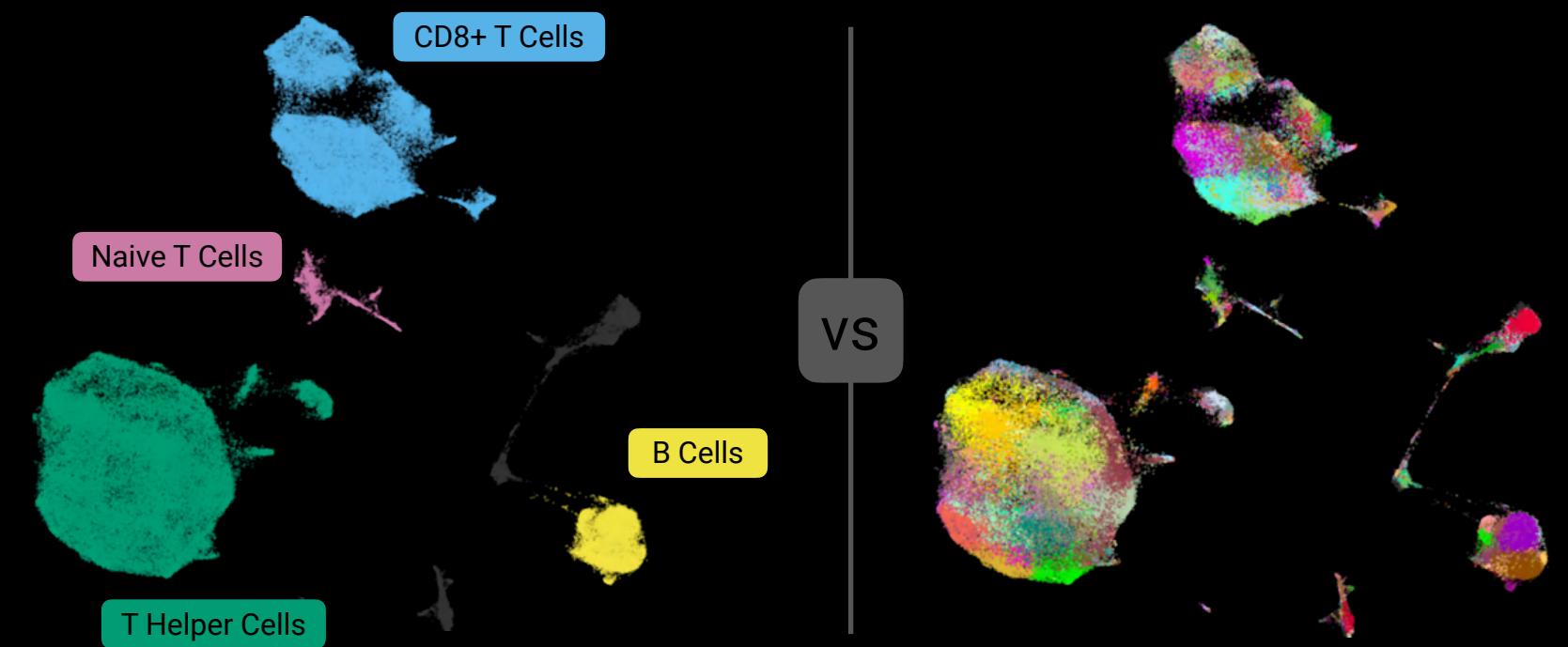


Visualization Challenges

Visualization Challenges

CLUSTER RESOLUTION

Focus on general or specific cellular phenotypes?



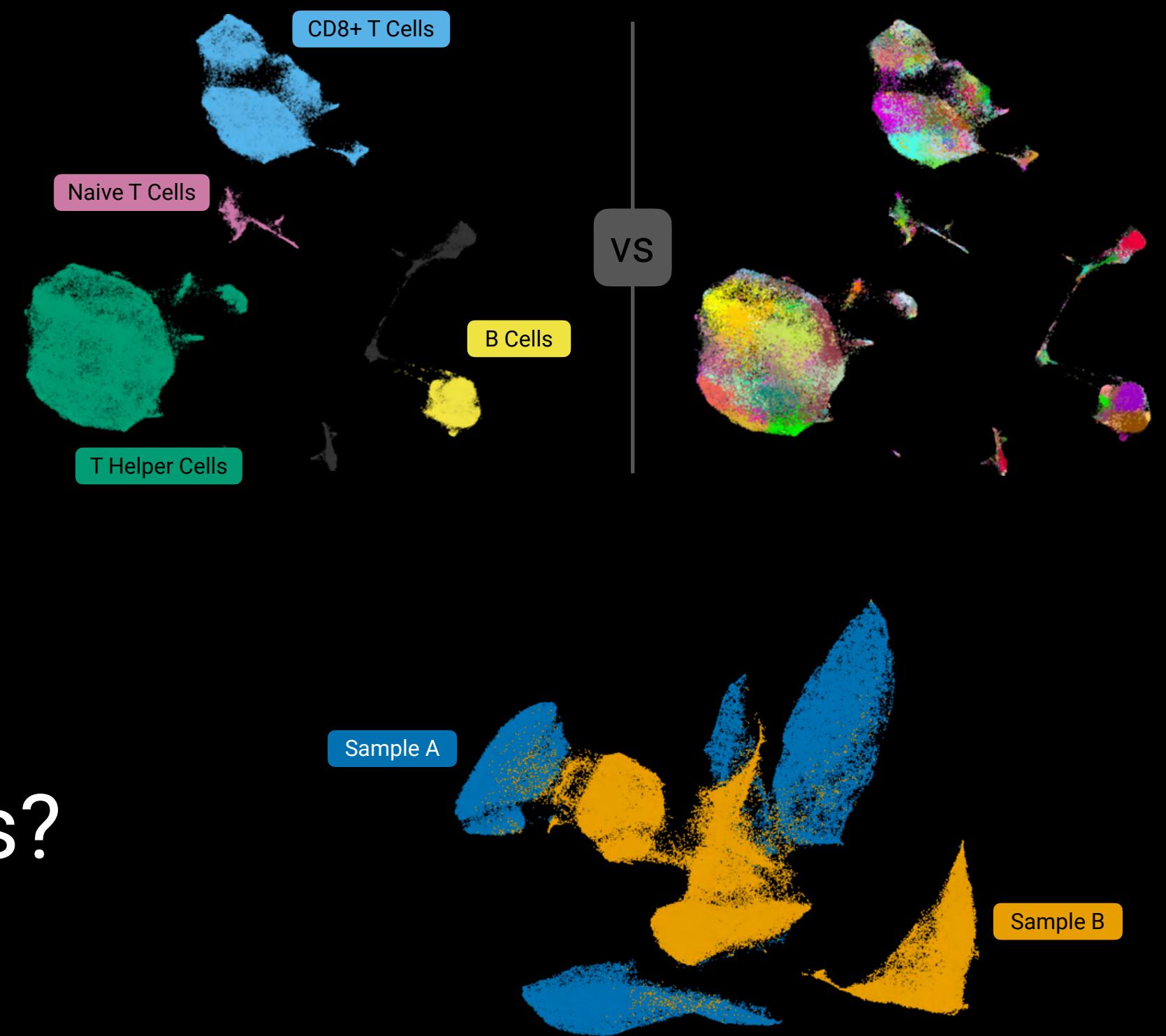
SAMPLE COMPARISON

How to handle batch effects and aligning embeddings?

Visualization Challenges

CLUSTER RESOLUTION

Focus on general or specific cellular phenotypes?



SAMPLE COMPARISON

How to handle batch effects and aligning embeddings?

EXPLORATION VS EXPLANATION

Is the visualization a representation of the clustering?

Annotate Expression Levels

FAUST Annotation + Clustering

ANNOTATE

Define expression levels

E.g.: Positive / Negative

→ Fully interpretable clusters



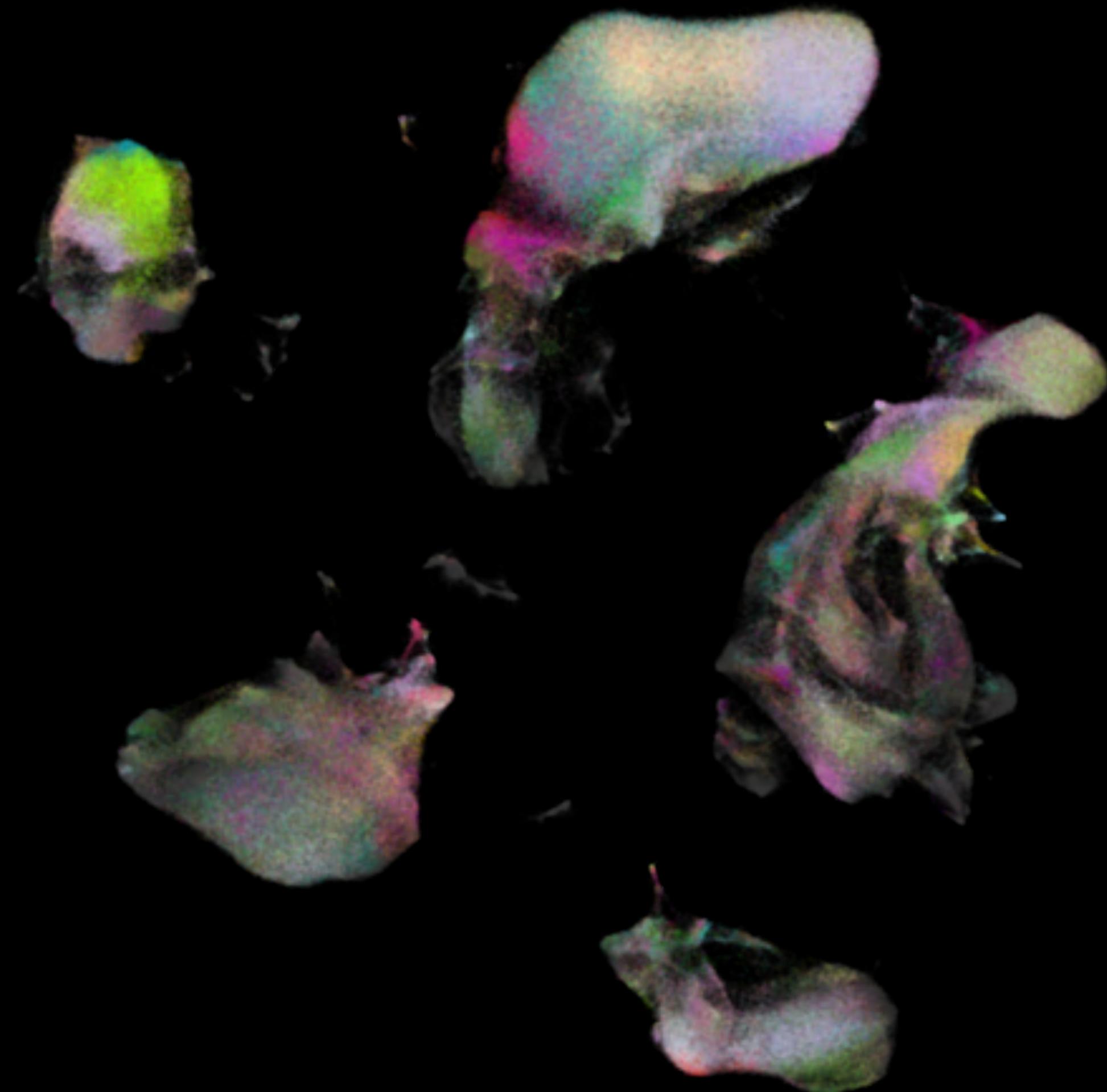
FAUST Annotation + Clustering

ANNOTATE

Define expression levels

E.g.: Positive / Negative

→ Fully interpretable clusters



Data from Subrahmanyam et al., 2018, *Journal for ImmunoTherapy of Cancer*.

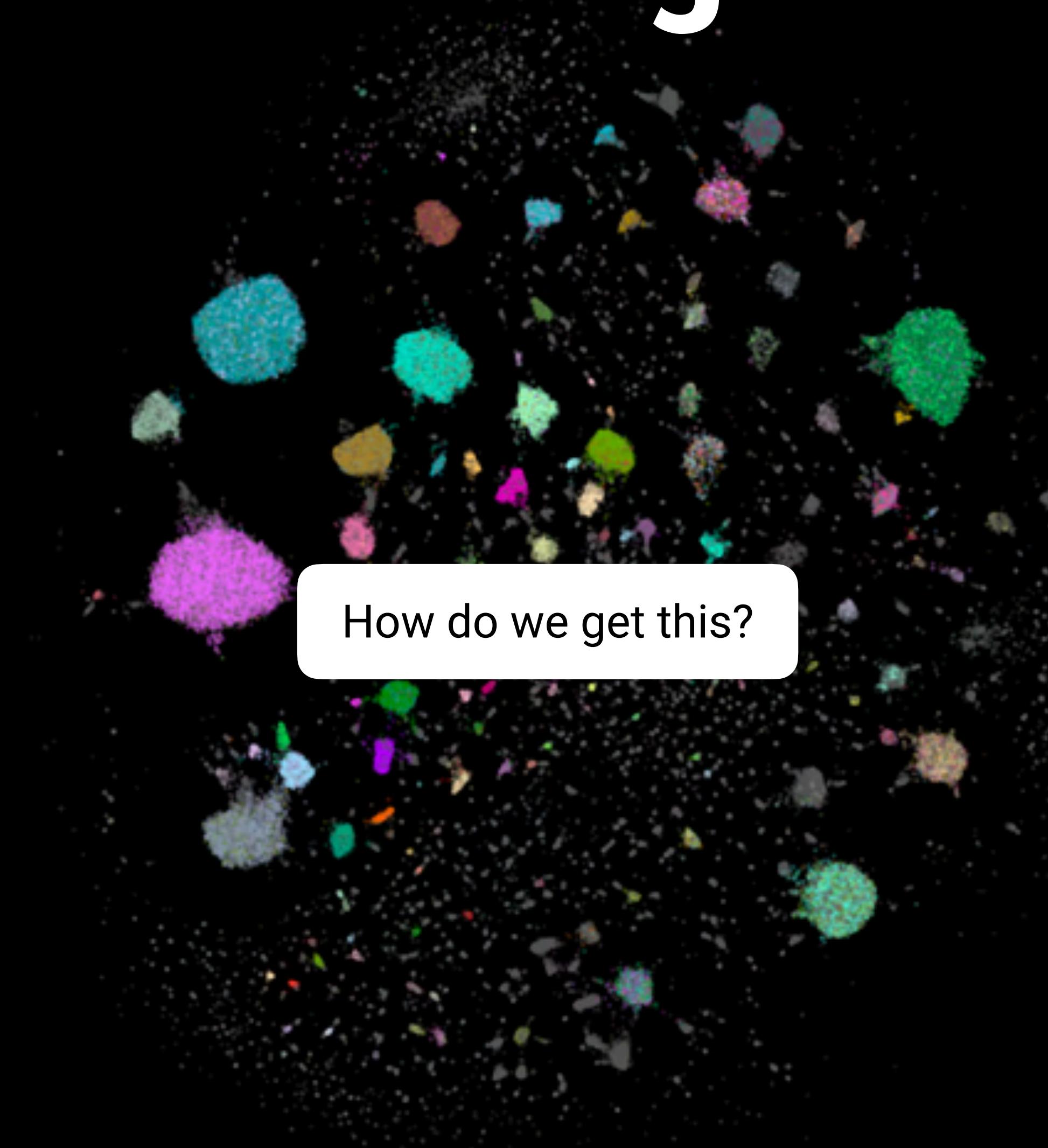
FAUST Annotation + Clustering

ANNOTATE

Define expression levels

E.g.: Positive / Negative

→ Fully interpretable clusters



Data from Subrahmanyam et al., 2018, *Journal for ImmunoTherapy of Cancer*.

Expression Annotation Transformation

Data Transformation

FOR EACH PHENOTYPE:

1. Remove outlier expression values
→ Winsorize to [1th, 99th] percentile
2. Remove inter marker differences
→ Normalize to zero mean and unit variance
3. Align marker expressions by their expression level
→ Translate mean to a fixed value

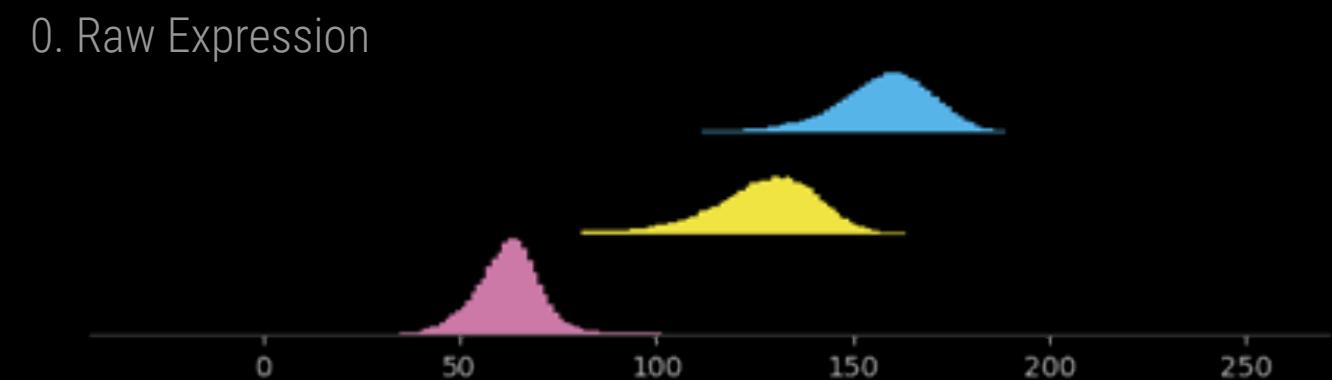
Data Transformation

CD3+ CD4+ CD8-

FOR EACH PHENOTYPE:

1. Remove outlier expression values

→ Winsorize to [1th, 99th] percentile



2. Remove inter marker differences

→ Normalize to zero mean and unit variance

3. Align marker expressions by their expression level

→ Translate mean to a fixed value

Data Transformation

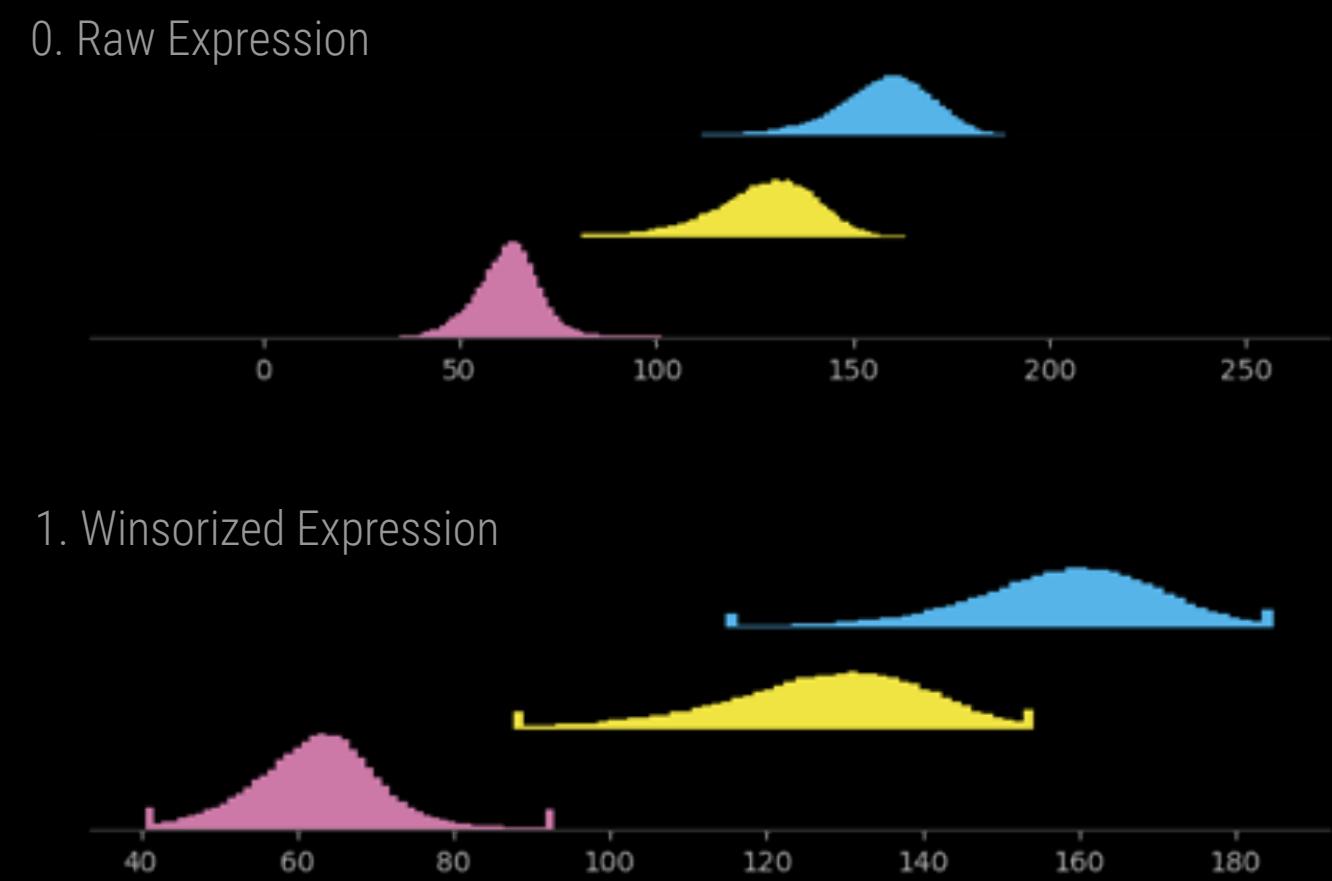
FOR EACH PHENOTYPE:

1. Remove outlier expression values
→ Winsorize to [1th, 99th] percentile

2. Remove inter marker differences
→ Normalize to zero mean and unit variance

3. Align marker expressions by their expression level
→ Translate mean to a fixed value

CD3+ CD4+ CD8-



Data Transformation

FOR EACH PHENOTYPE:

1. Remove outlier expression values

→ Winsorize to [1th, 99th] percentile

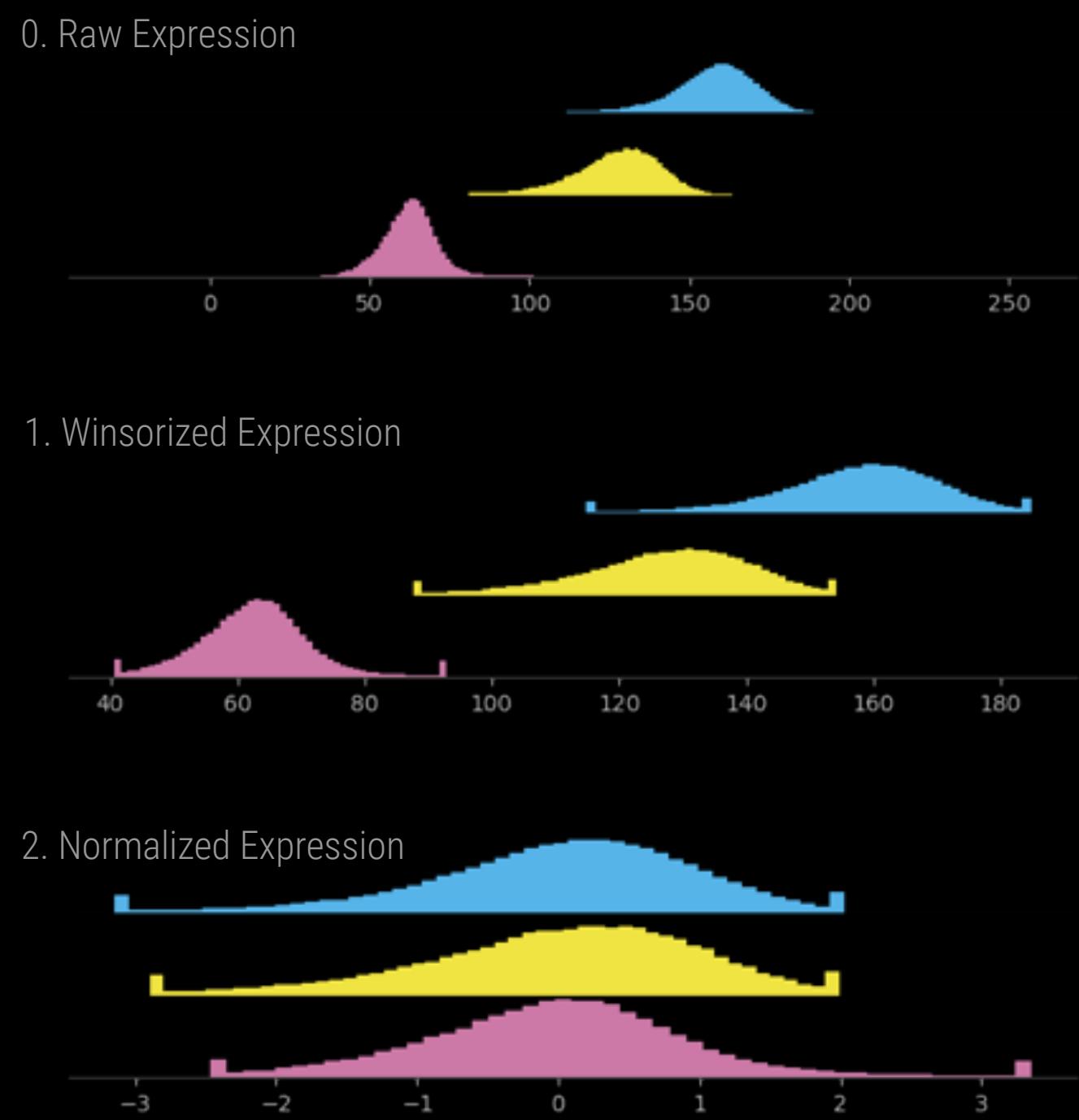
2. Remove inter marker differences

→ Normalize to zero mean and unit variance

3. Align marker expressions by their expression level

→ Translate mean to a fixed value

CD3+ CD4+ CD8-



Data Transformation

FOR EACH PHENOTYPE:

1. Remove outlier expression values

→ Winsorize to [1th, 99th] percentile

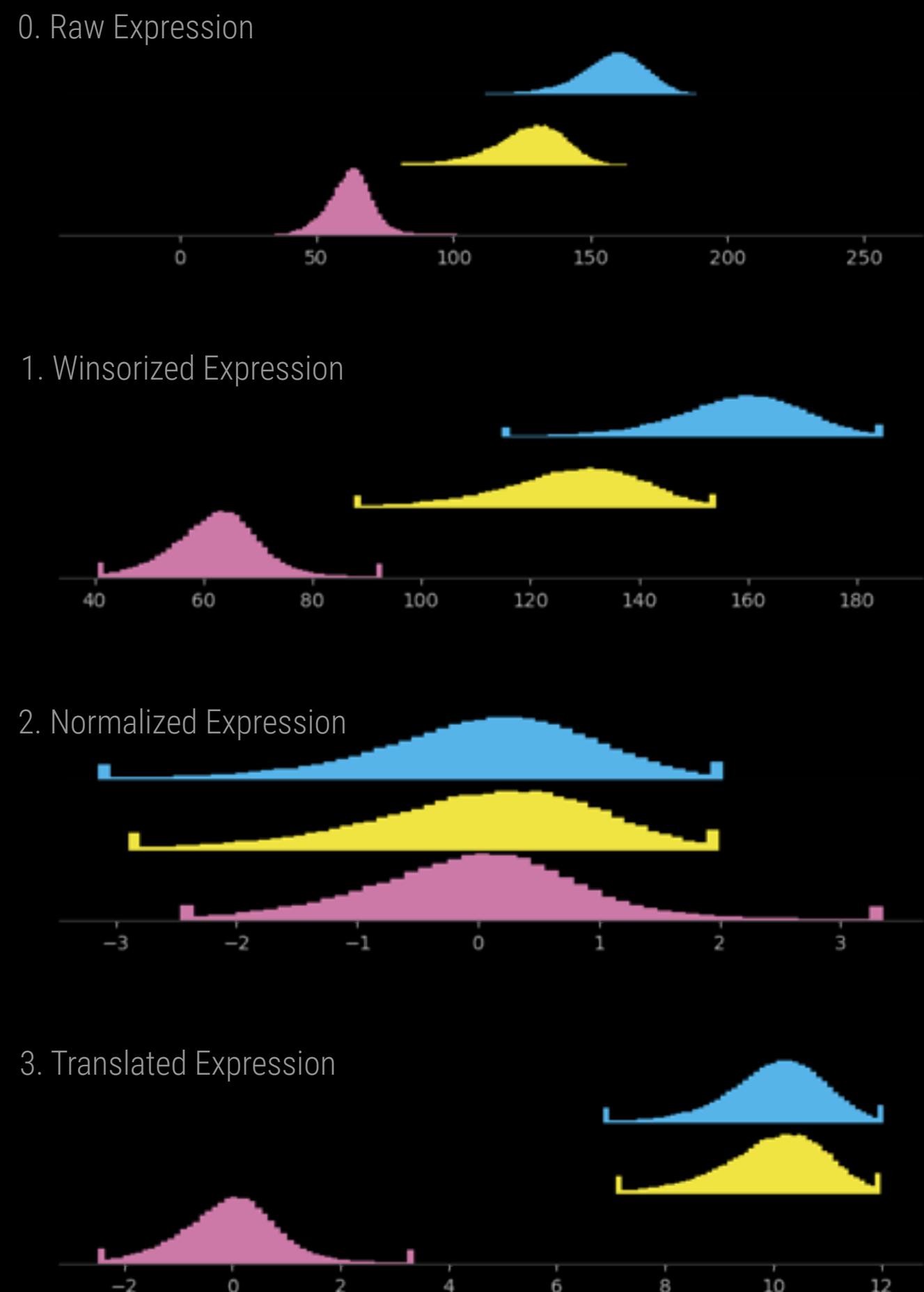
2. Remove inter marker differences

→ Normalize to zero mean and unit variance

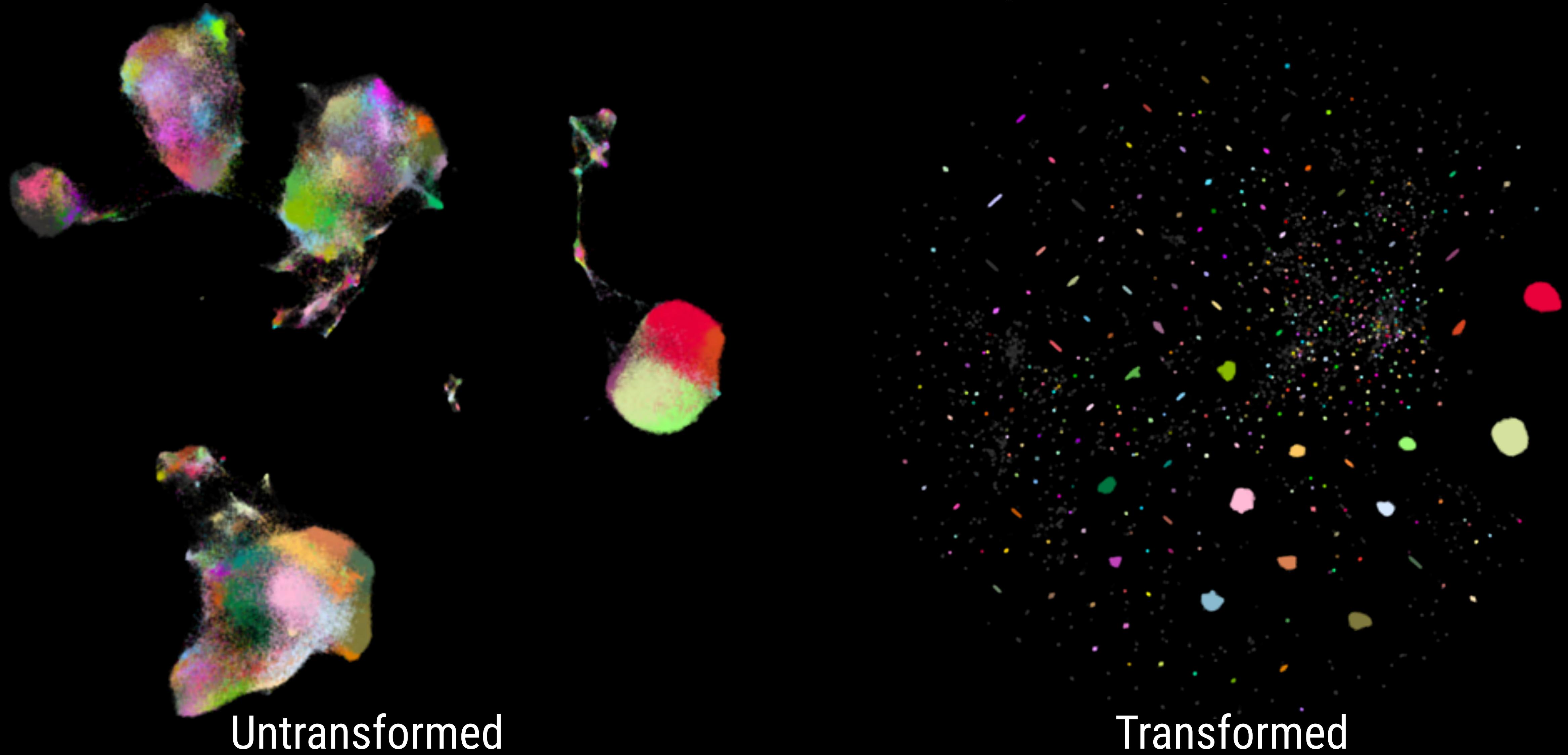
3. Align marker expressions by their expression level

→ Translate mean to a fixed value

CD3+ CD4+ CD8-

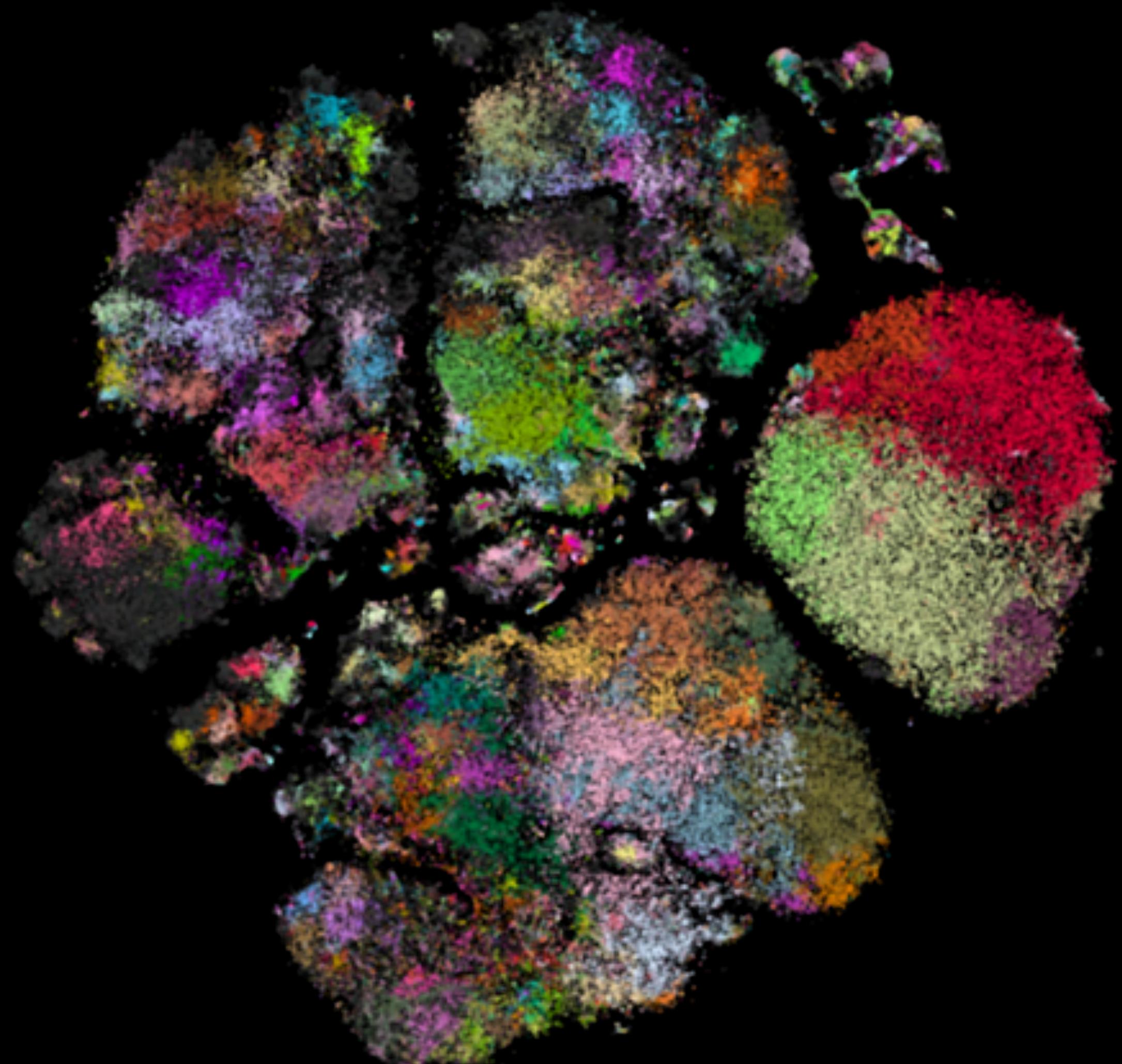


UMAP Embedding



Tumor sample 6 from Mair et al., 2022, *Nature*.

t-SNE Embedding



Untransformed



Transformed

Tumor sample 6 from Mair et al., 2022, *Nature*.

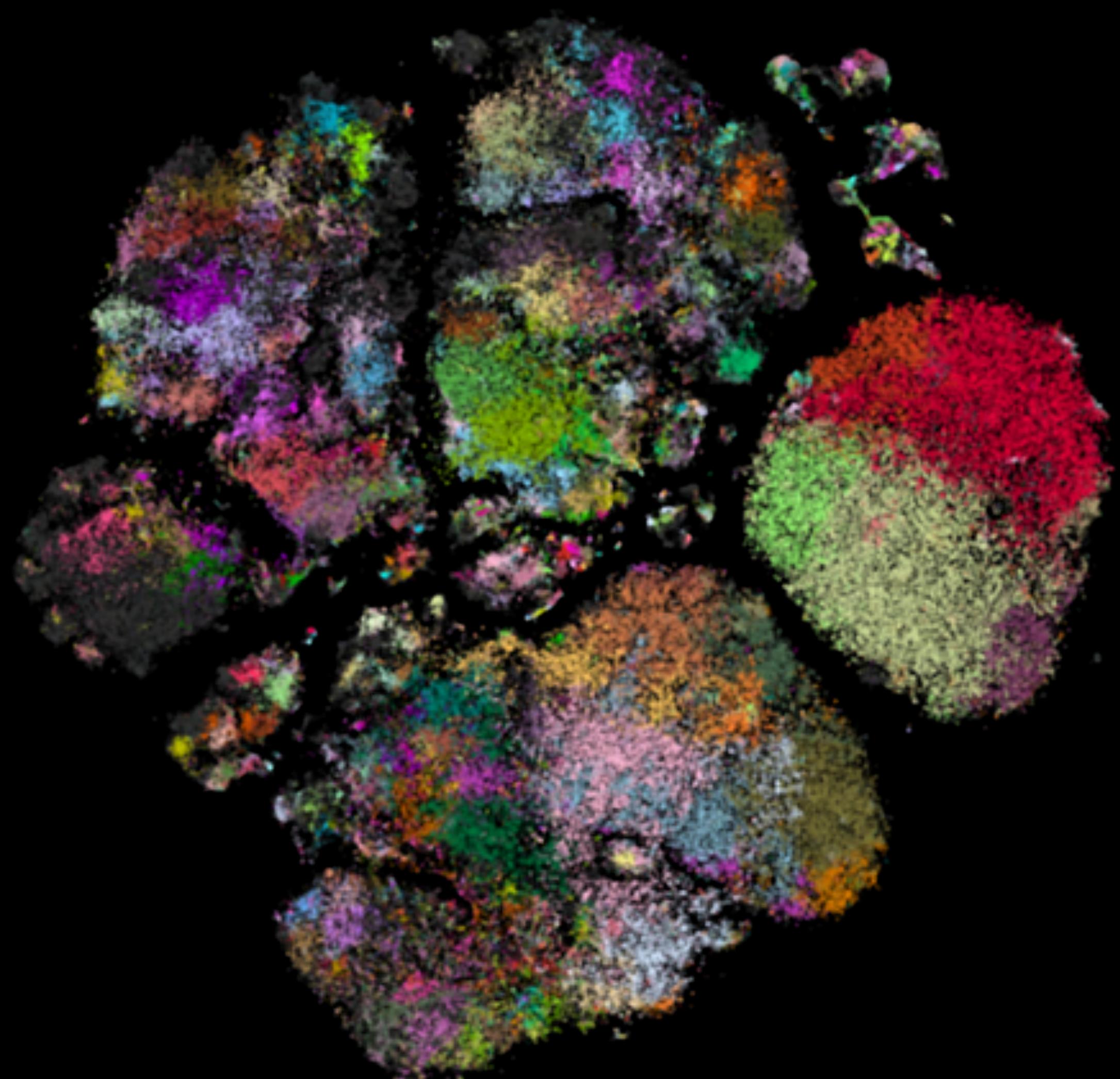
VAE Embedding

Untransformed

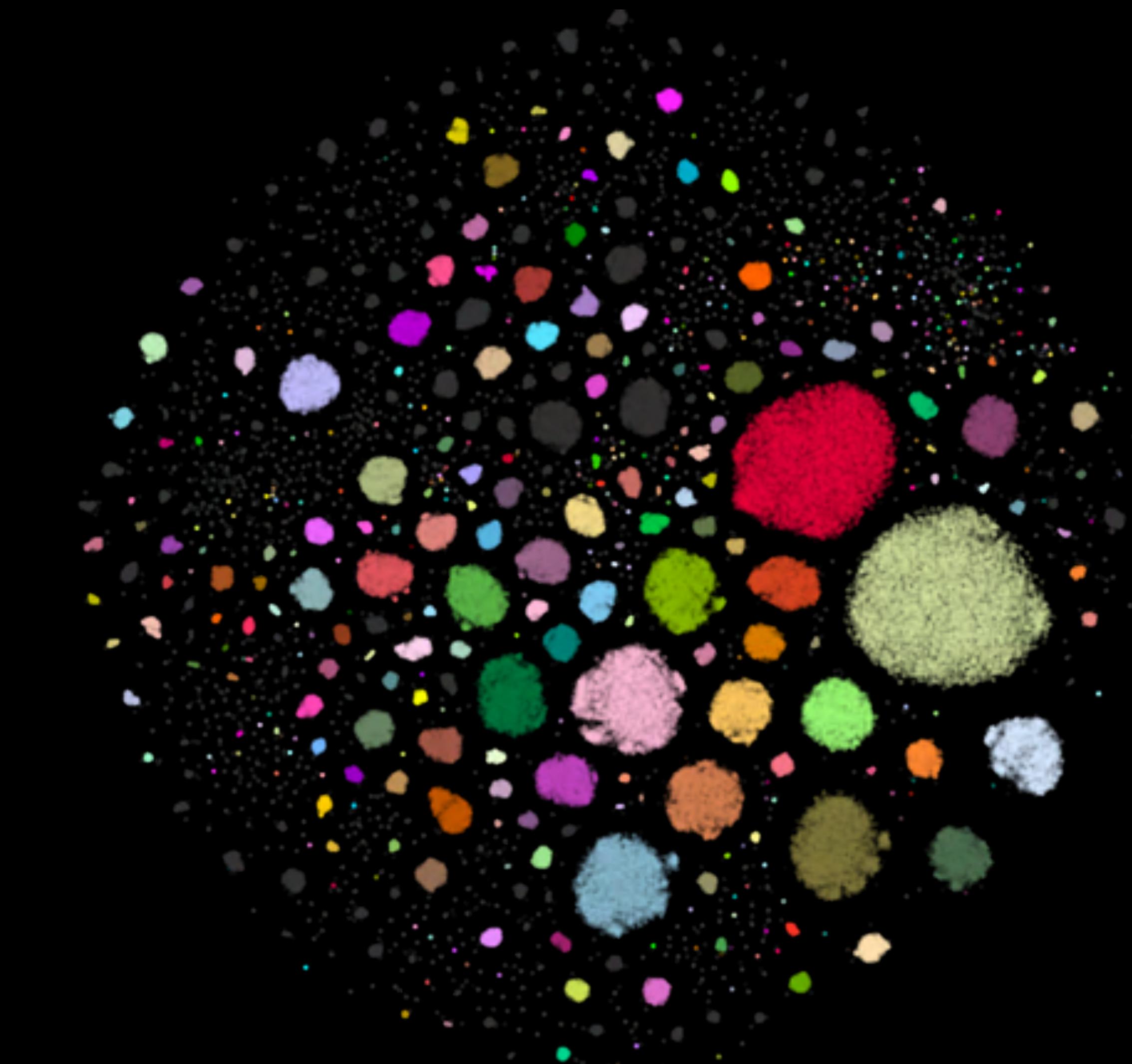
Transformed

Tumor sample 6 from Mair et al., 2022, *Nature*.

Cluster Coherence



Untransformed

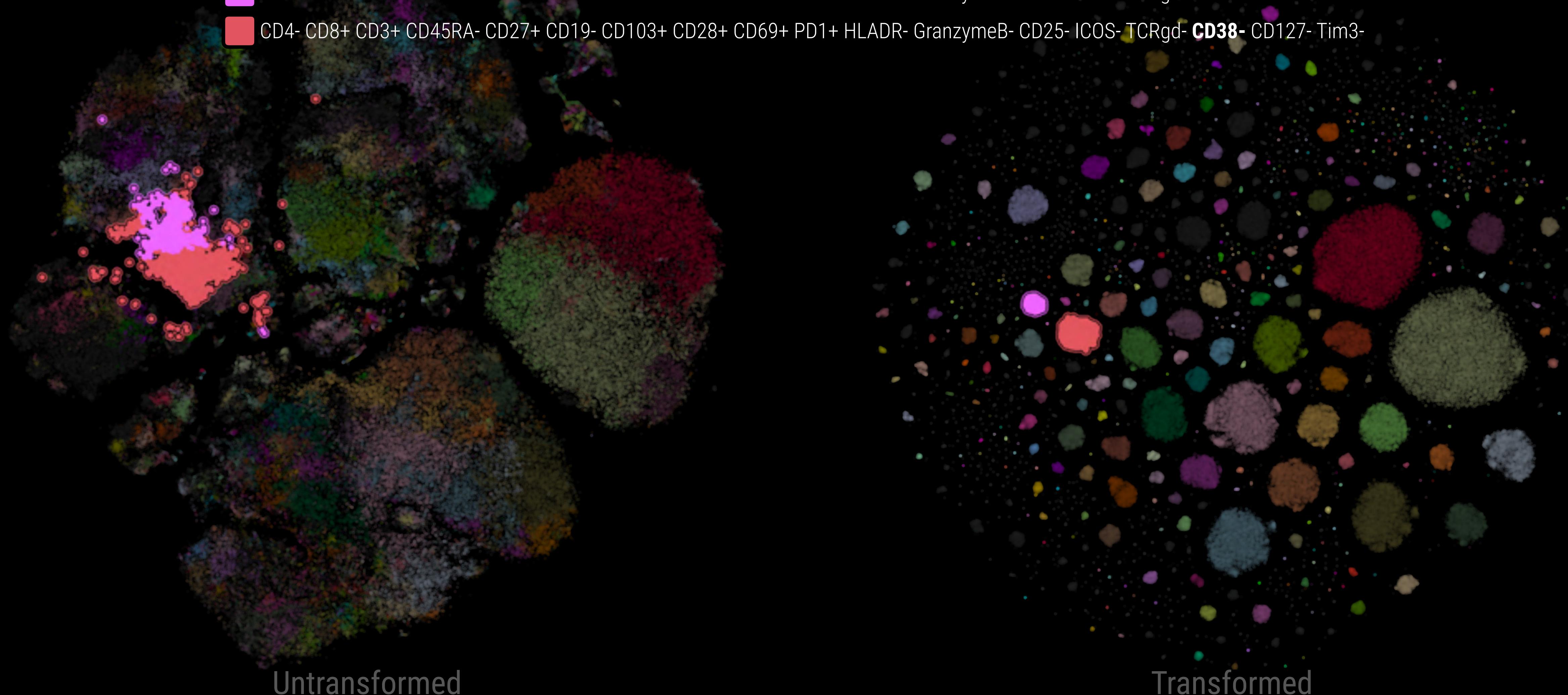


Transformed

Tumor sample 6 from Mair et al., 2022, *Nature*.

CD38 Expression Difference

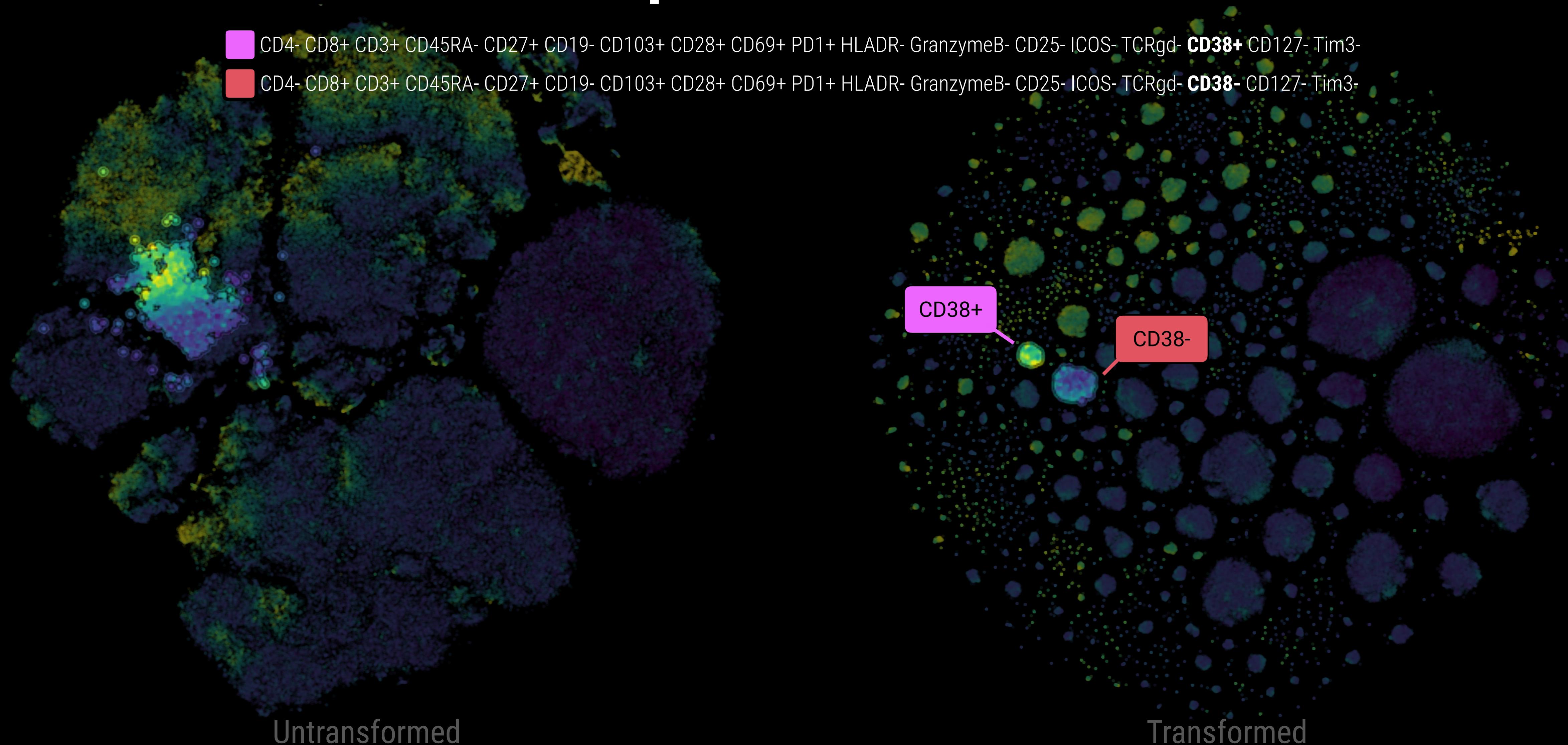
■ CD4- CD8+ CD3+ CD45RA- CD27+ CD19- CD103+ CD28+ CD69+ PD1+ HLADR- GranzymeB- CD25- ICOS- TCRgd- **CD38+** CD127- Tim3-
■ CD4- CD8+ CD3+ CD45RA- CD27+ CD19- CD103+ CD28+ CD69+ PD1+ HLADR- GranzymeB- CD25- ICOS- TCRgd- **CD38-** CD127- Tim3-



Tumor sample 6 from Mair et al., 2022, *Nature*.

CD38 Expression Difference

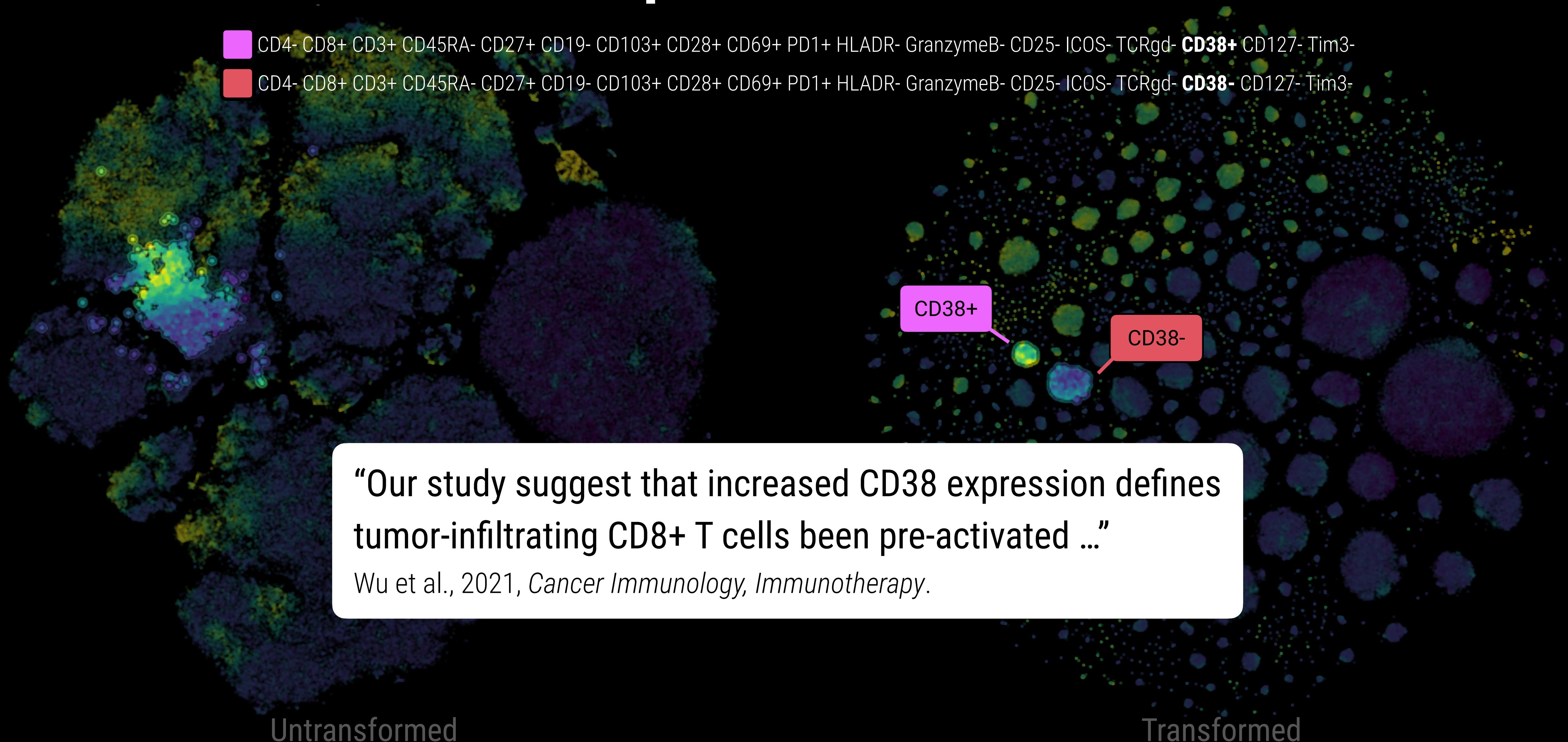
■ CD4- CD8+ CD3+ CD45RA- CD27+ CD19- CD103+ CD28+ CD69+ PD1+ HLADR- GranzymeB- CD25- ICOS- TCRgd- **CD38+** CD127- Tim3-
■ CD4- CD8+ CD3+ CD45RA- CD27+ CD19- CD103+ CD28+ CD69+ PD1+ HLADR- GranzymeB- CD25- ICOS- TCRgd- **CD38-** CD127- Tim3-



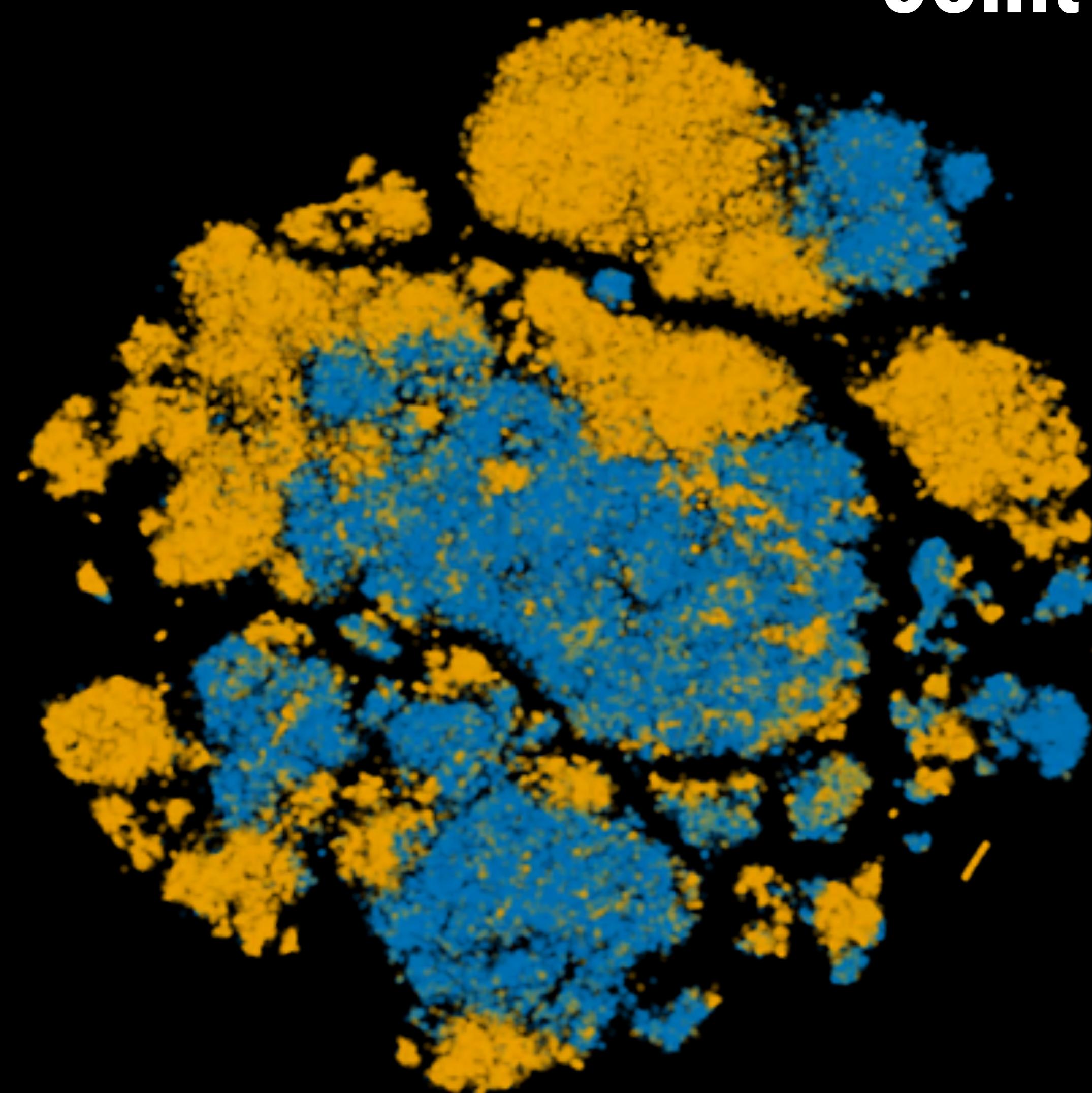
Tumor sample 6 from Mair et al., 2022, *Nature*.

CD38 Expression Difference

■ CD4- CD8+ CD3+ CD45RA- CD27+ CD19- CD103+ CD28+ CD69+ PD1+ HLADR- GranzymeB- CD25- ICOS- TCRgd- **CD38+** CD127- Tim3-
■ CD4- CD8+ CD3+ CD45RA- CD27+ CD19- CD103+ CD28+ CD69+ PD1+ HLADR- GranzymeB- CD25- ICOS- TCRgd- **CD38-** CD127- Tim3-



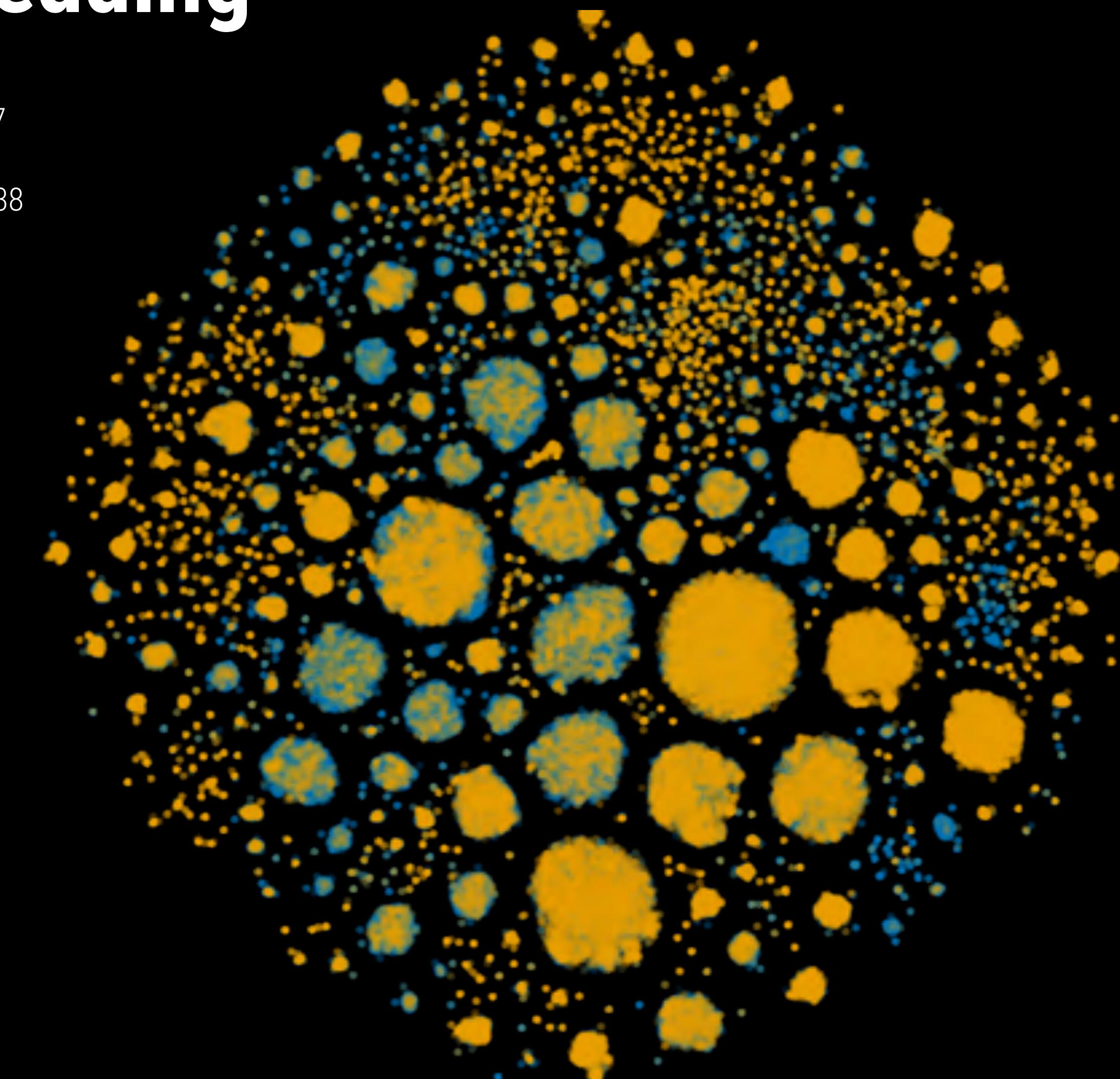
Joint Embedding



Untransformed

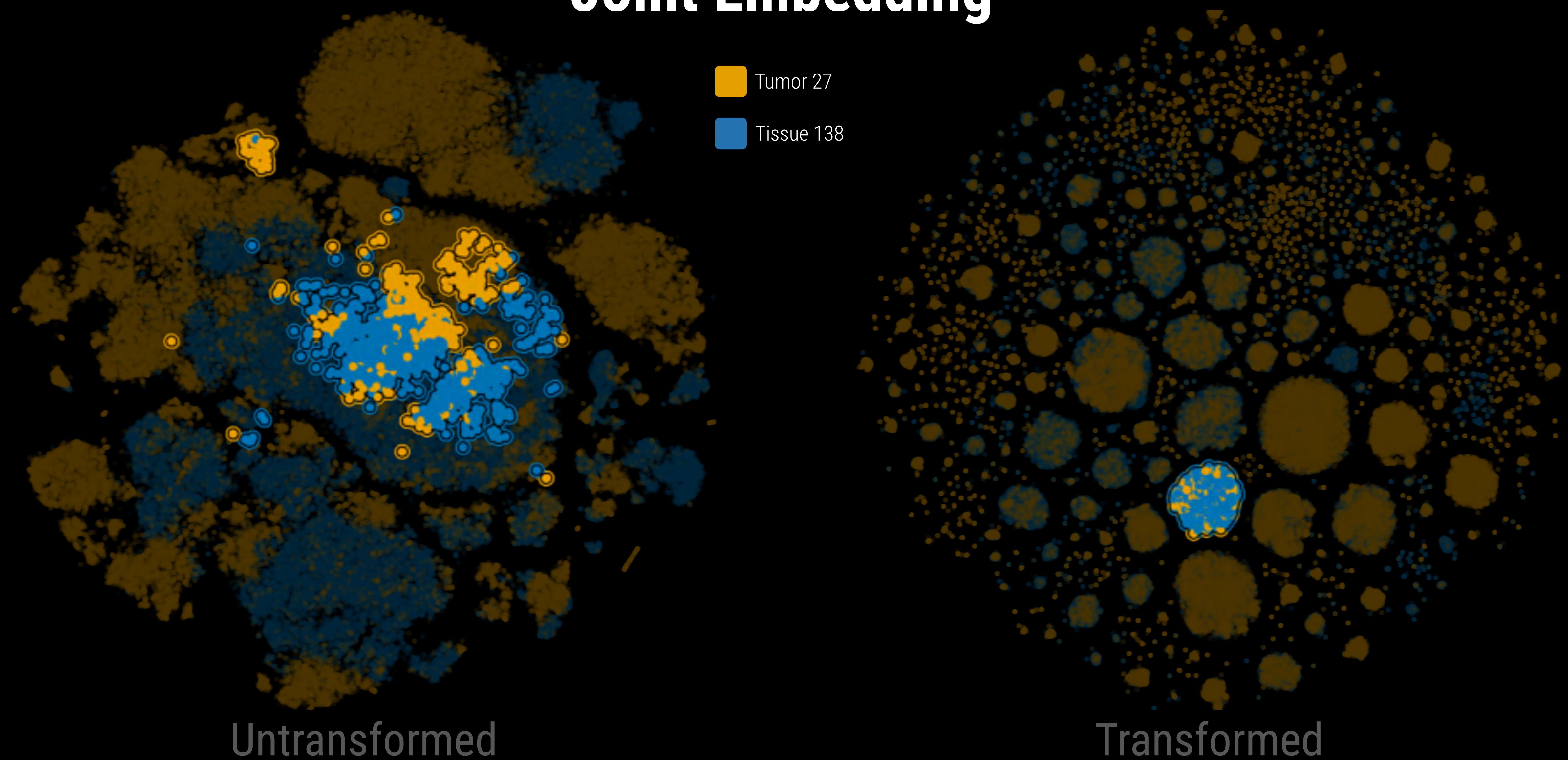
Tumor 27

Tissue 138



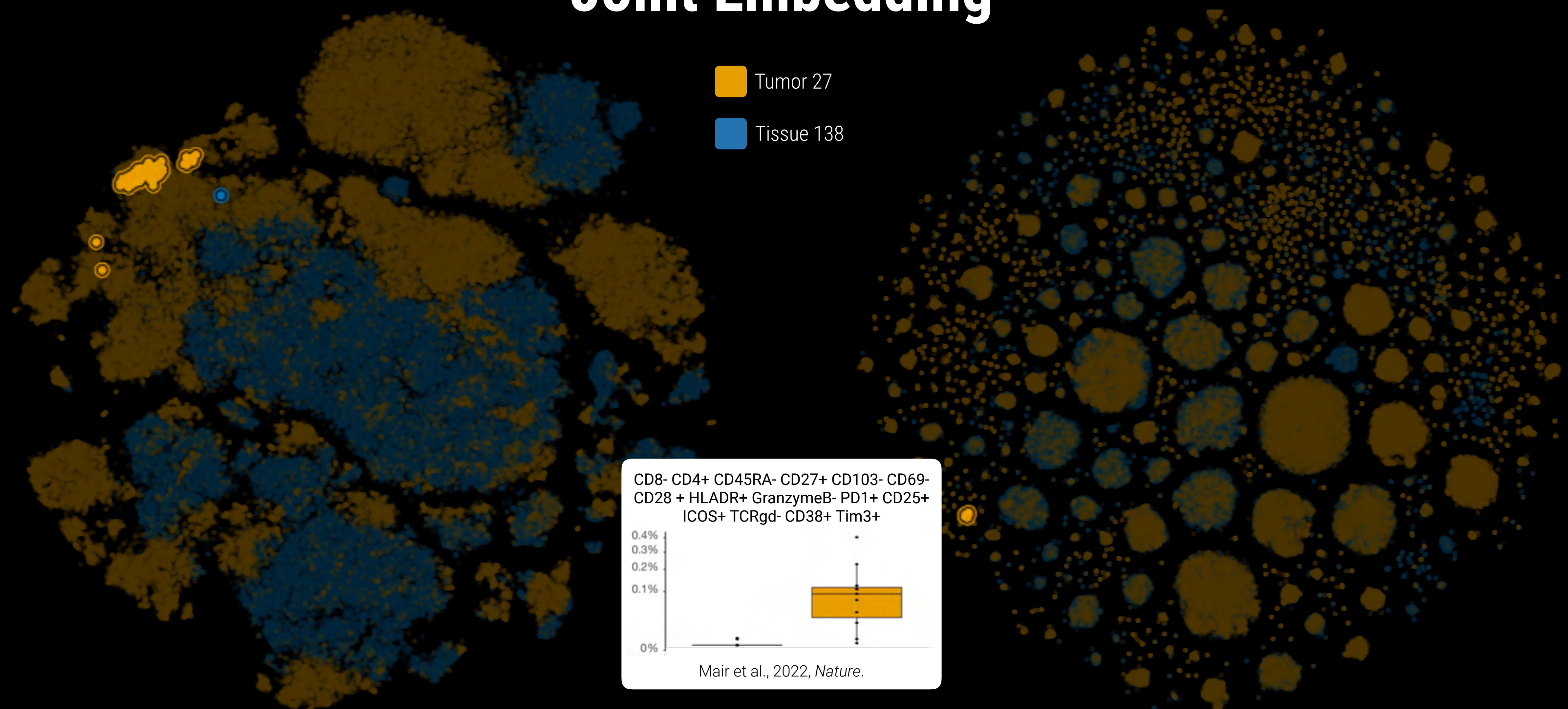
Transformed

Joint Embedding



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

Joint Embedding



Untransformed

Transformed

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

CONCLUSION

- “Tune” the data
- Use a data transformation close to your objective
- The annotation transformation is not bound to FAUST

CONCLUSION

- “Tune” the data and the embedding method
- Use a data transformation close to your objective
- The annotation transformation is not bound to FAUST

Thank You!

CODE

<https://github.com/flekschas-ozette/ismb-biovis-2022>

<https://github.com/RGLab/faust>

PAPER

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

CONTACT



Fritz@ozette.com



Evan@ozette.com



Malisa@ozette.com

👋 See You at our Poster R-008

👉 Related MLCSB Talk

Today at 12:20 in Room "Madison CD"