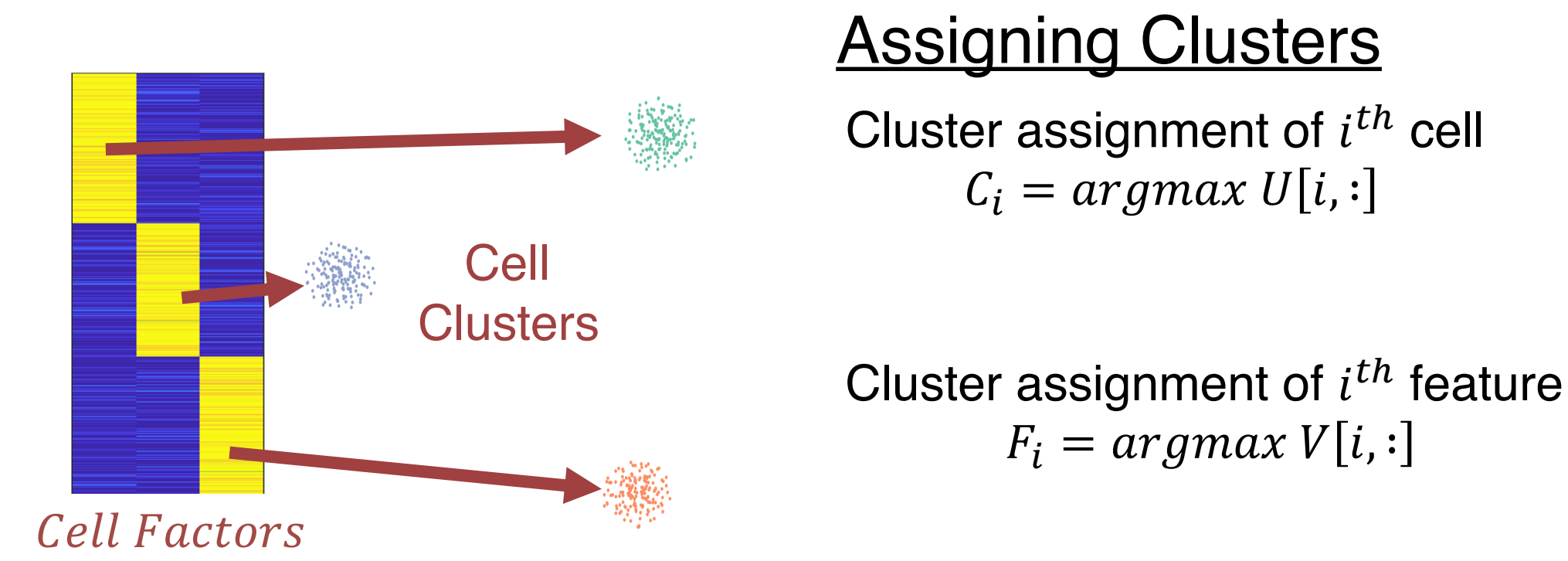
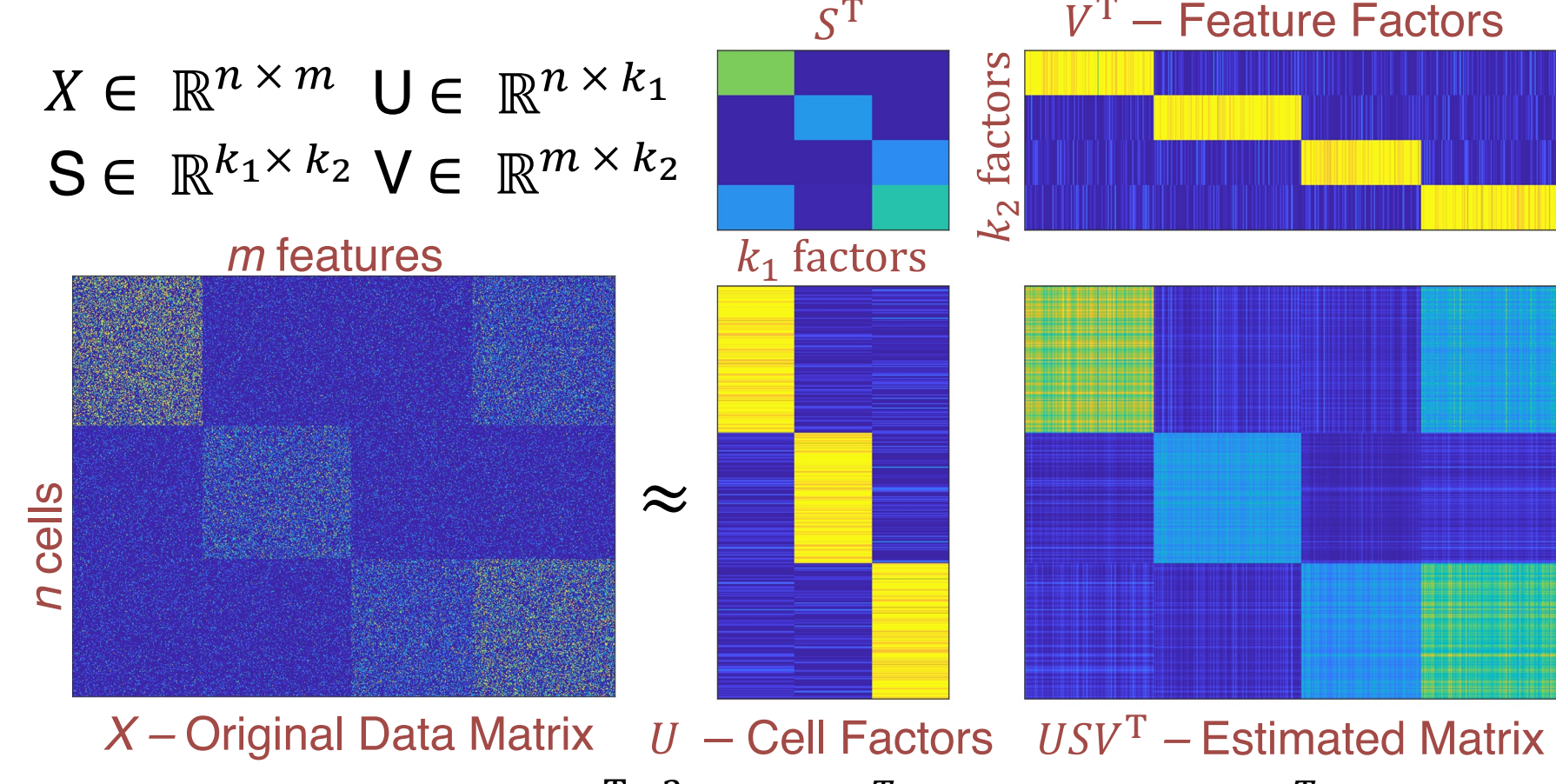


## Introduction

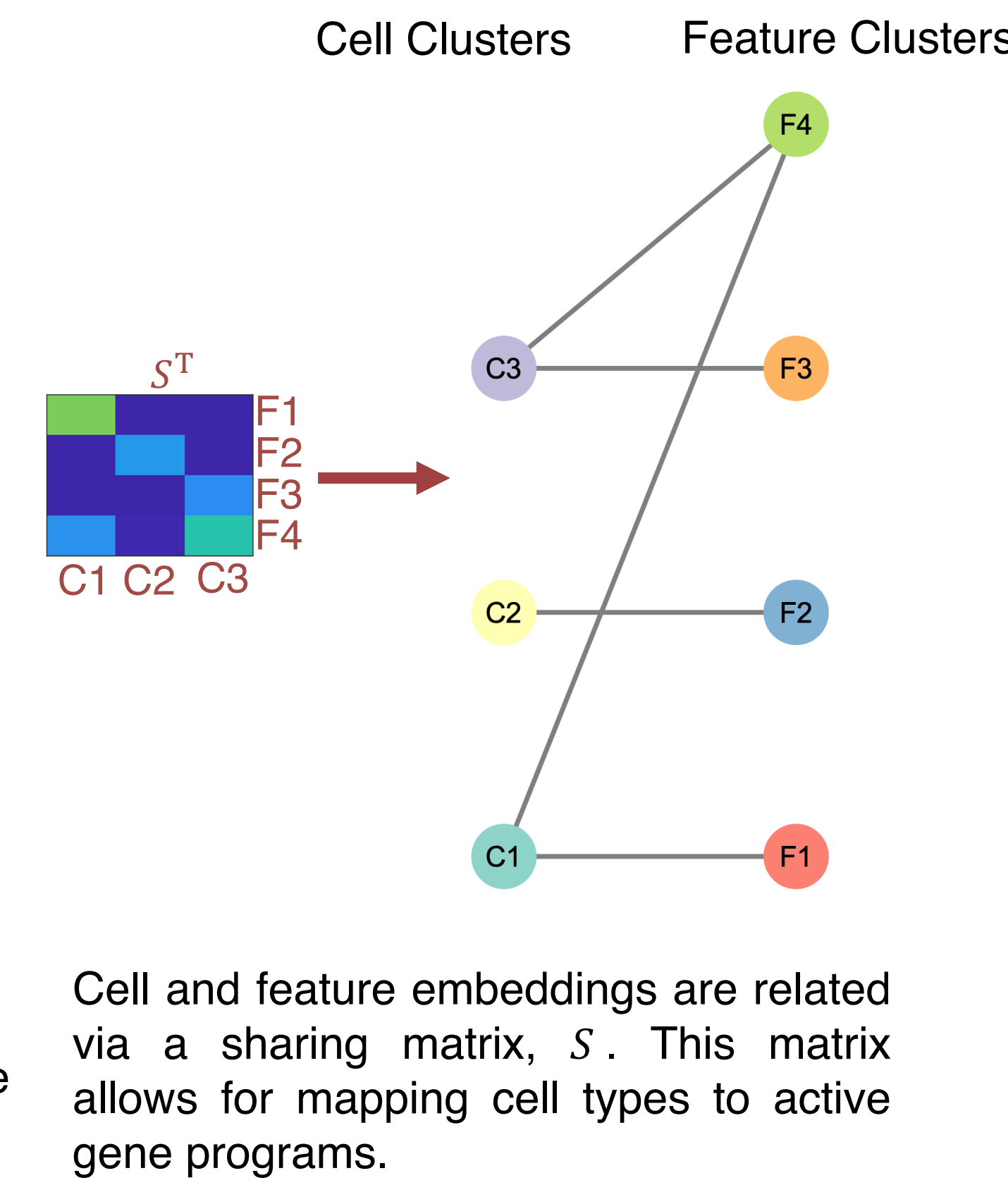
- Single-cell omic technologies, such as single cell RNA-sequencing (scRNA-seq), captures high-dimensional molecular profiles, such as gene expression at the individual cell level.
- Dimensionality reduction is a powerful strategy for understanding single cell data. A common workflow includes clustering and annotation of cell types and states. Existing packages for this task use dimensionality reduction approaches such as principal components analysis (PCA) and non-negative matrix factorization (NMF).
- We apply a different dimensionality reduction algorithm, orthogonal non-negative matrix tri-factorization (O-NMTF), that enables us to define the low-dimensional embedding of cells and genomic features simultaneously.
- We use this lower-dimensional embeddings for clustering of single cell datasets and identifying active gene programs.

## Methods

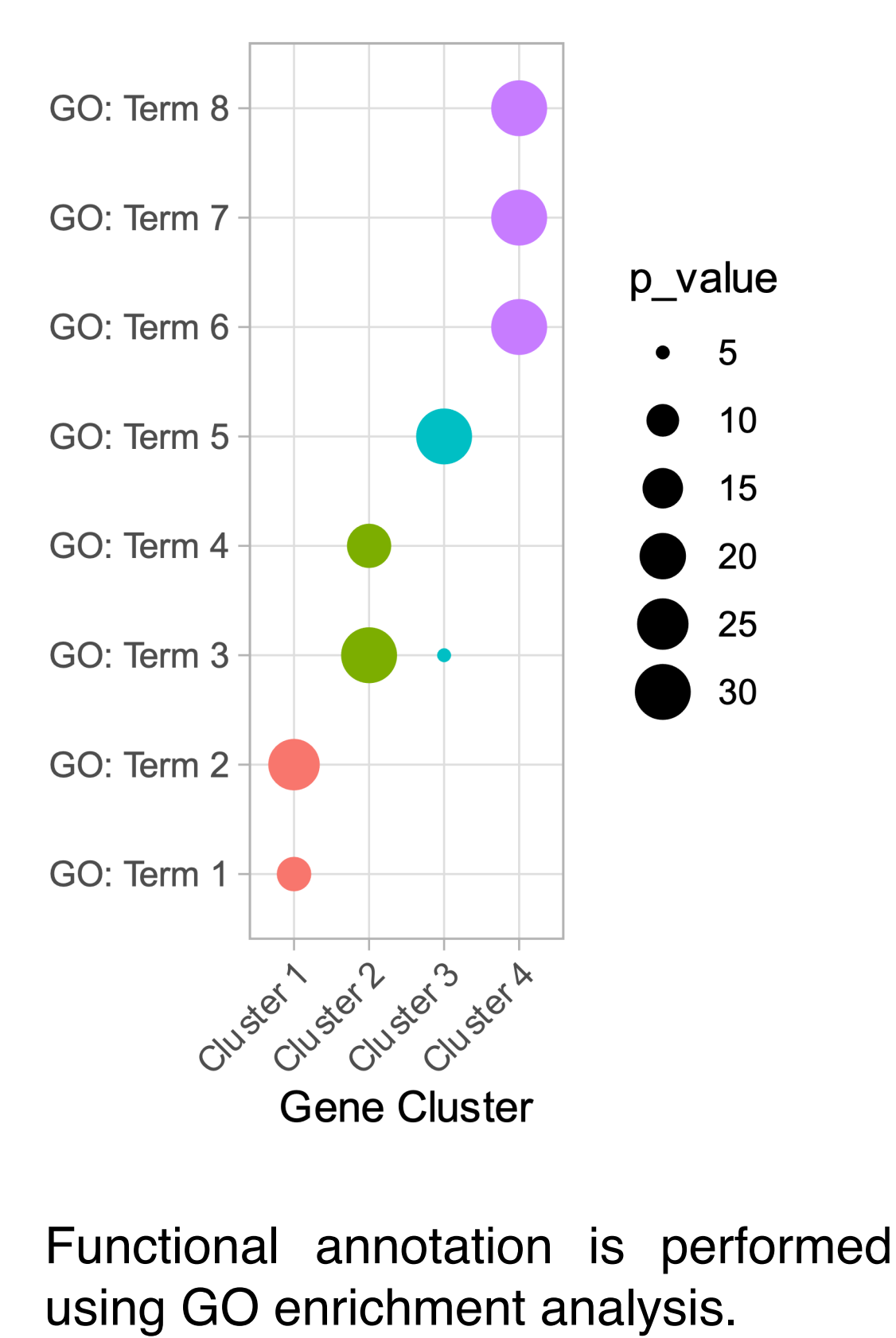
### Orthogonal Non-Negative Matrix Tri-Factorization



### Relate Cell and Feature Clusters



### Feature Interpretation



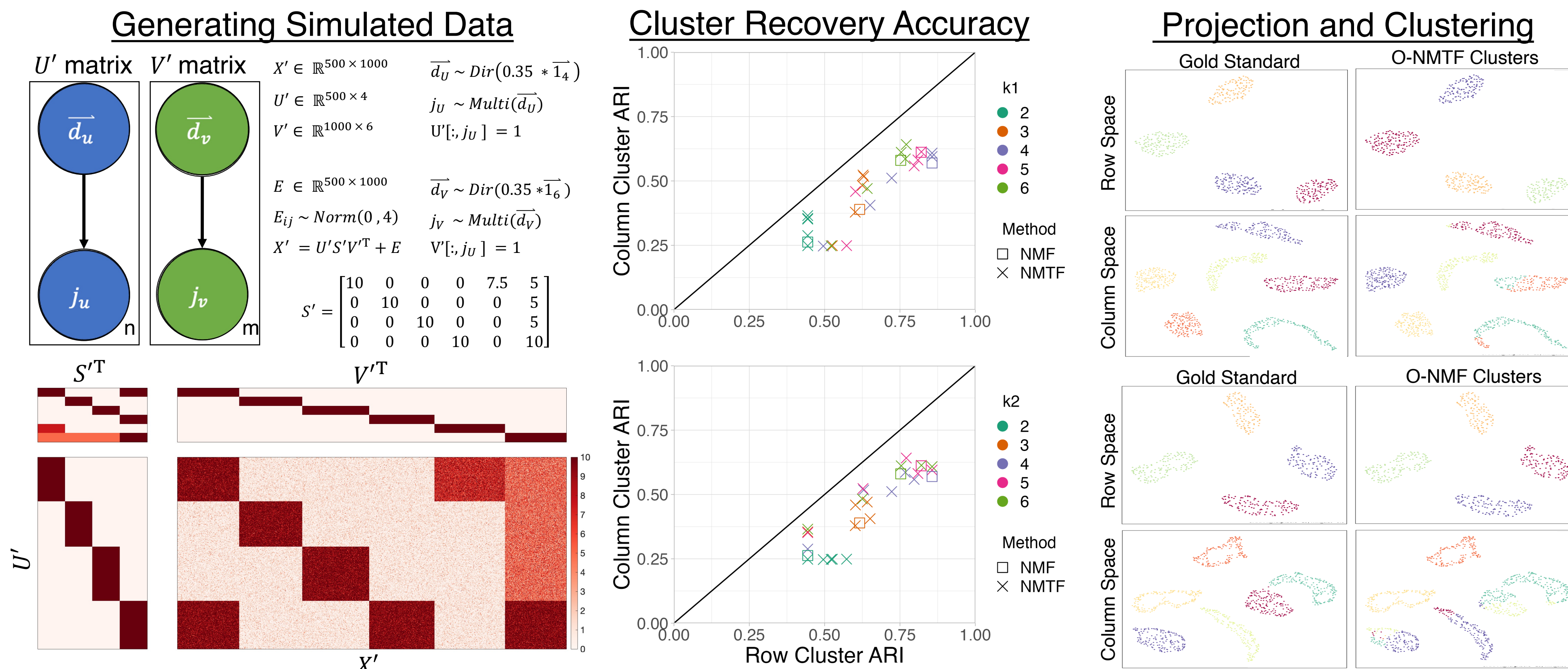
## Datasets

- 1. Simulated Matrices:**
  - Pre-defined cell and feature clusters.
  - 500 cells, 1000 features.
- 2. Human Peripheral Blood Mononuclear Cells (PBMC<sup>1</sup>)**
  - Labeled cell types.
  - 2638 cells, 6566 genes
- 3. 2D and 3D retinal organoid system during differentiation**
  - Uncharacterized cell types.
  - 94,065 cells, 22904 genes.

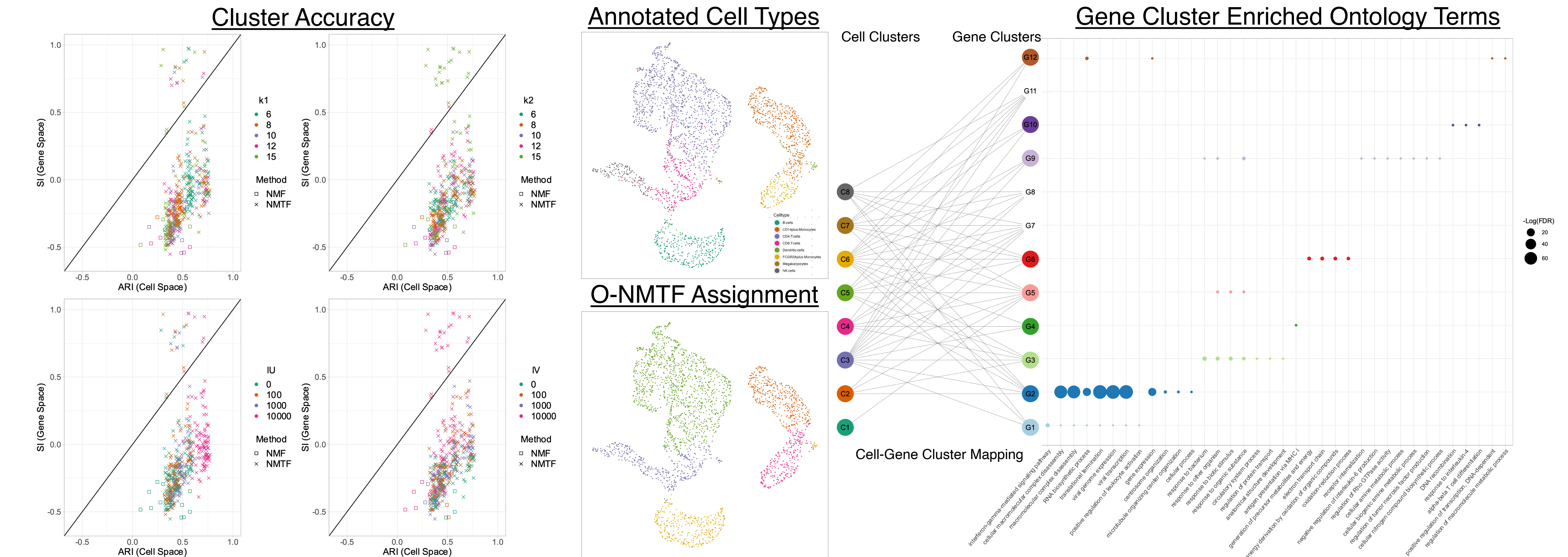
## Clustering Methods

- Orthogonal non-negative matrix tri-factorization (O-NMTF): Finds orthogonal cell and feature factors. Provides a bi-clustering of the "omic" matrix. Cell and feature clusters are related via a relation matrix,  $S$ , that allows for many-to-many relationships.
- Orthogonal non-negative matrix factorization (O-NMF): Finds feature clusters and orthogonal cell clusters. Provides a bi-clustering of the "omic" matrix. Cell and features are related to-one<sup>[2,3]</sup>.

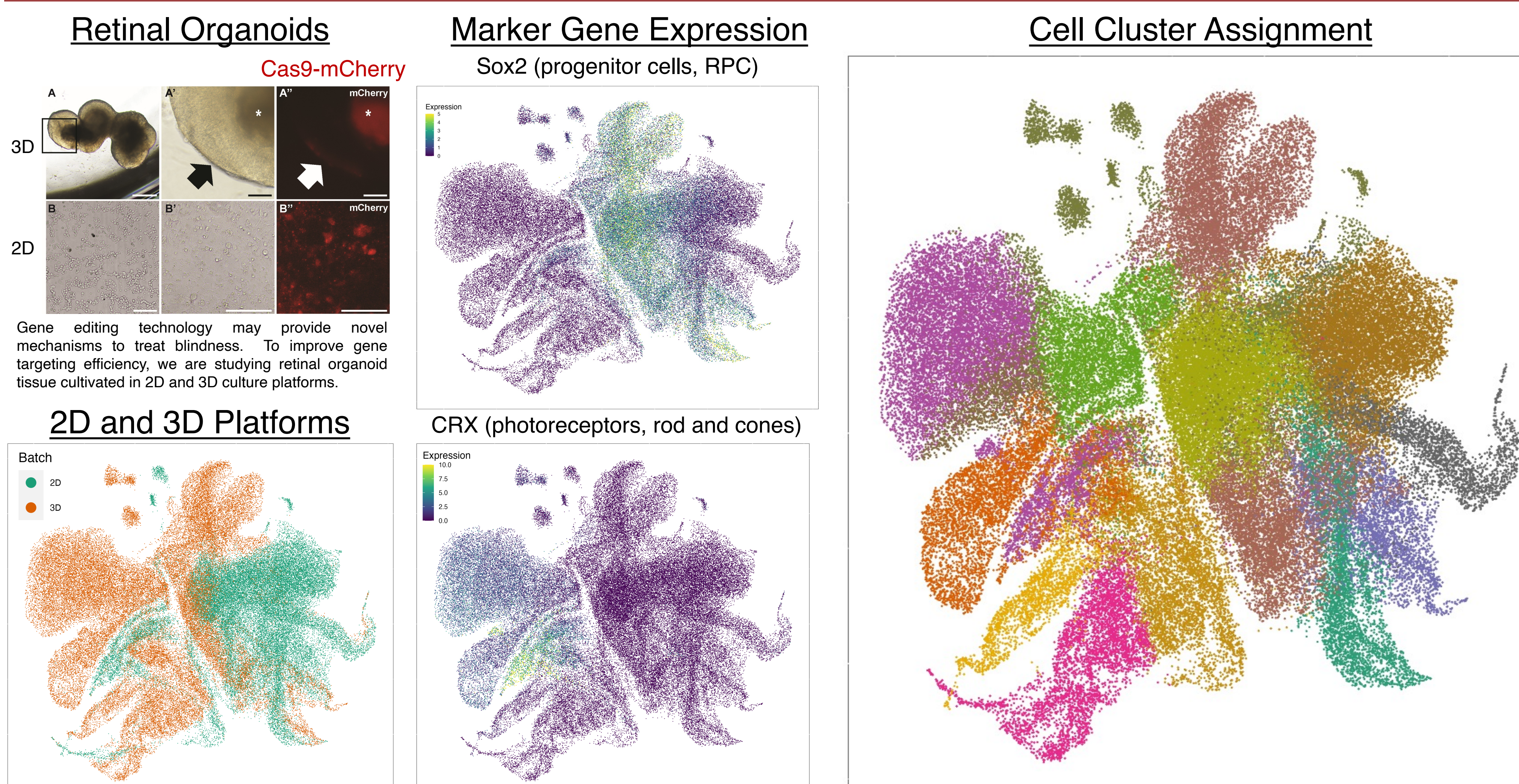
## Result 1: O-NMTF simultaneously recovers row and column clusters better than O-NMF



## Result 2: O-NMTF outperforms O-NMF at recovering accurate cell types and meaningful gene clusters for PBMC data



## Result 3: O-NMTF identifies platform specific cell clusters in a 2D and 3D retinal organoid system



## Conclusion

### Summary

- We developed orthogonal non-negative matrix factorization (O-NMTF), a novel approach for clustering and annotating single cell omic data.
- O-NMTF embeds the cell and feature space into two lower dimensional spaces. Optimization over these hyperparameters allows O-NMTF to bi-cluster data more accurately than standard O-NMF on simulated data.
- We evaluated the performance of O-NMTF on a single cell RNA-seq PBMC dataset. O-NMTF recovered known cell labels more accurately than O-NMF.
- O-NMTF bi-clusters the cell and features of a single cell omic experiment. The clusters are related by a relationship matrix. We demonstrate that these features clusters portray biologically relevant information using GO term enrichment and cell marker enrichment. In both the PBMC, and retinal organoids, cluster specific features are used to annotate cell clusters.

### References

- [1] Buenrostro J. D. *et al. Cell* 173, 1535-1548.e16 (2018)
- [2] Davarajan K. *PLoS Computational Biology* 4(7), (2008)
- [3] Lee D. D. and Sueng, S., *Advances in Neural Information Processing Systems* 13, 556-562, MIT Press, (2001)

### Acknowledgements

This research is supported by National Science Foundation IOS Grant #2010789, National Institute of Health Grant #8U01EY032333-03, and National Institute of Health Grant #5R01GM117339-05.