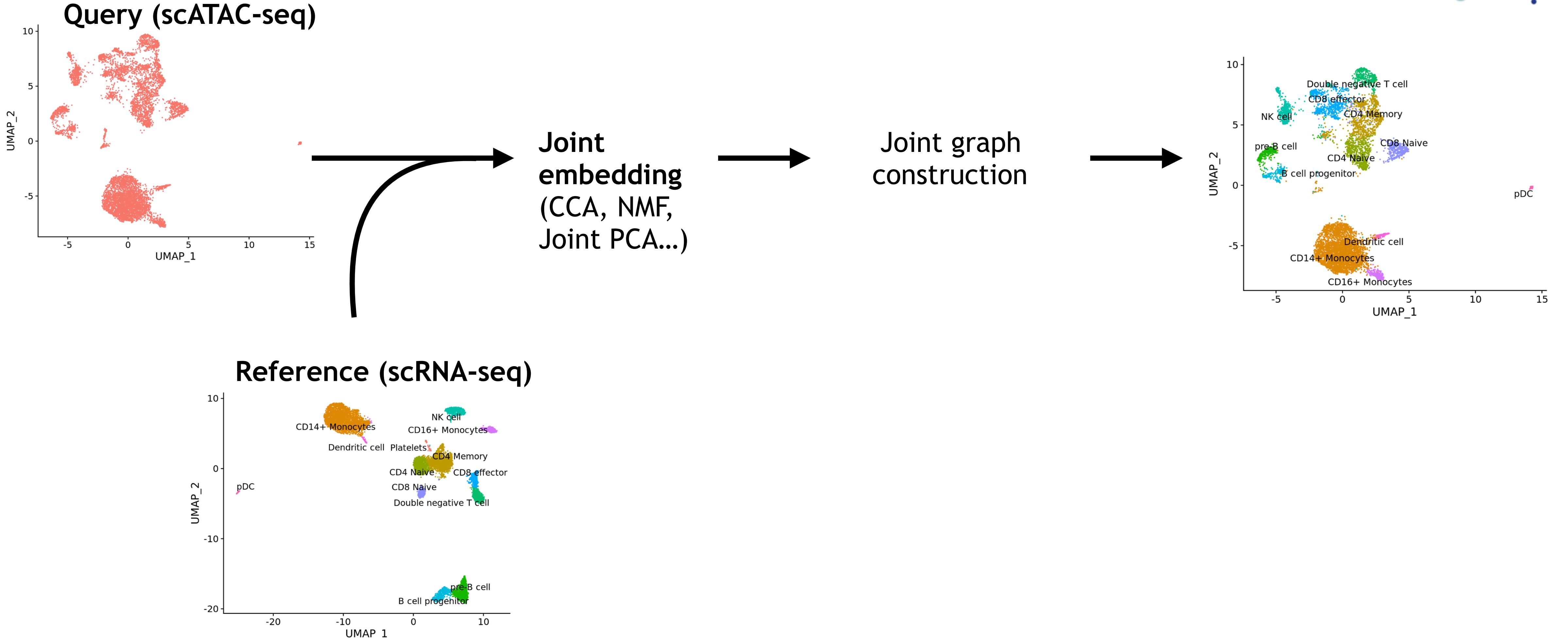




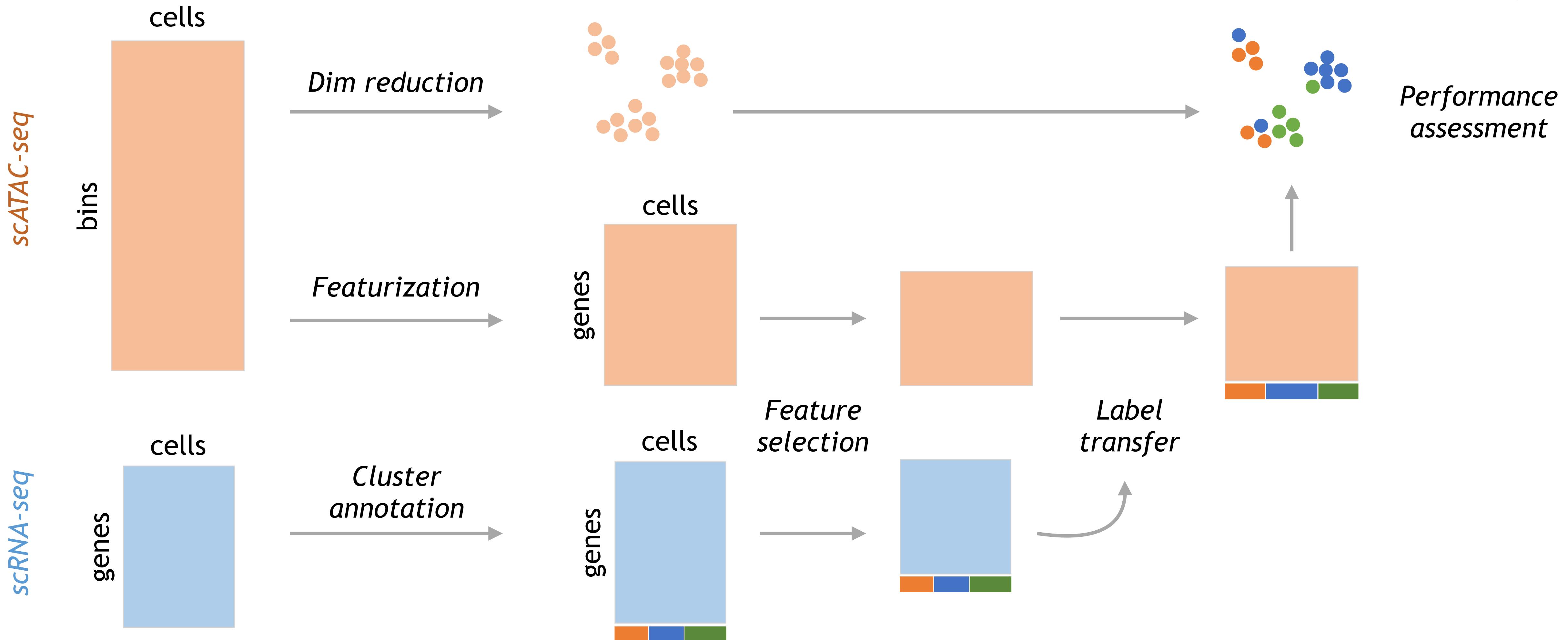
# Optimizing label propagation between scRNA-seq and scATAC-seq

Emma Dann

# Testing methods for label propagation



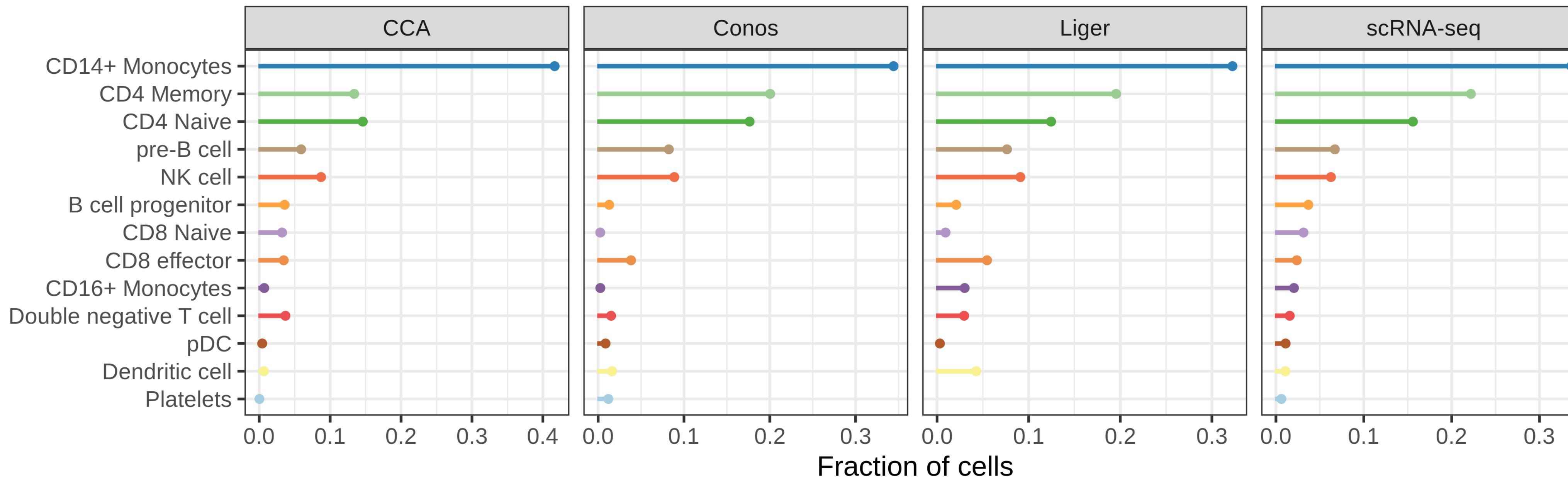
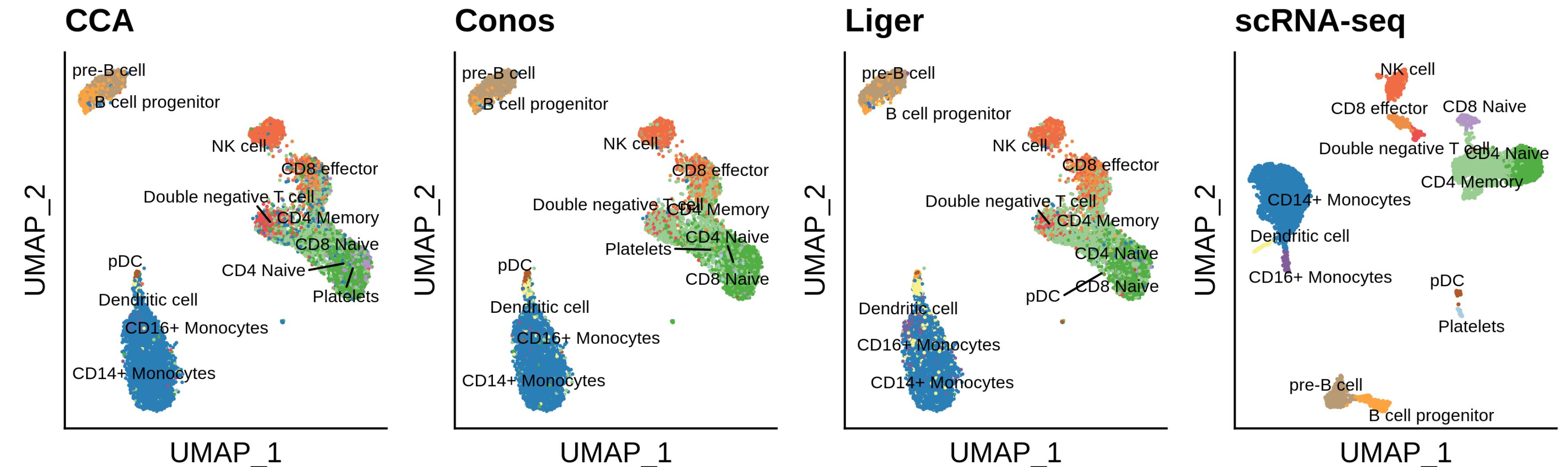
# Label transfer workflow



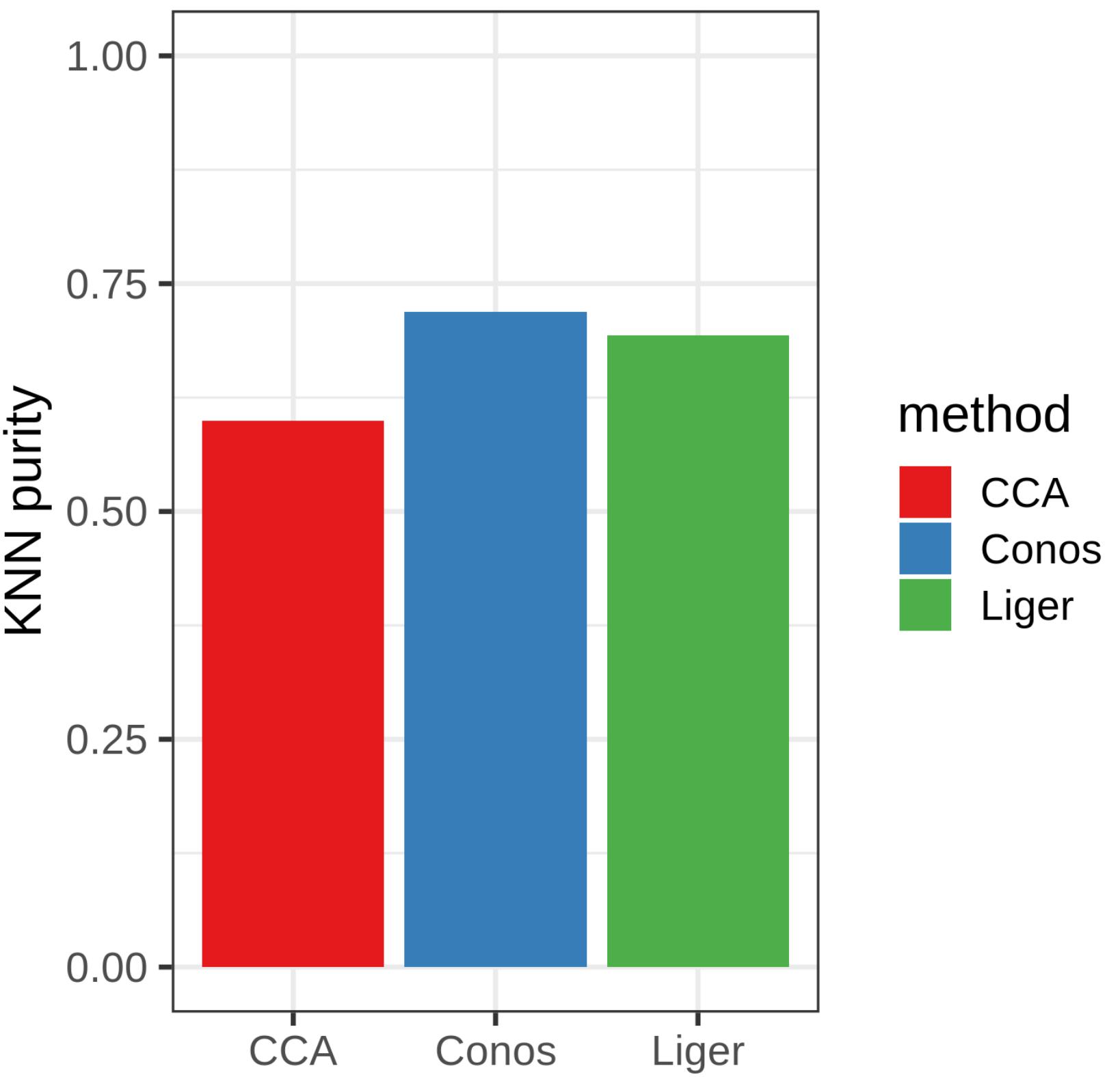
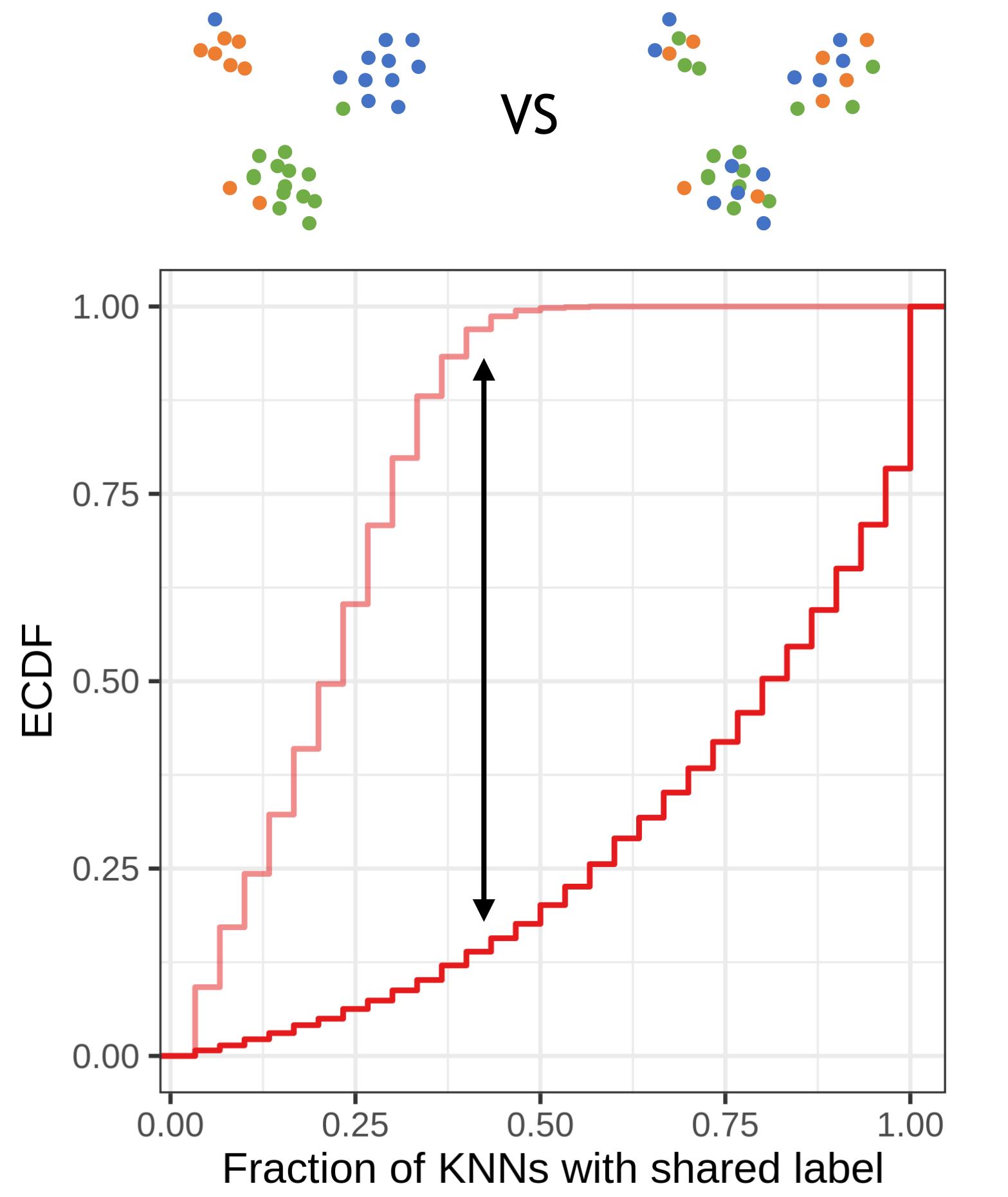
# Tested integration methods

Method	Reference	Model for joint embedding	Label/feature propagation	Reason for excluding
Seurat CCA	Stuart et al. Cell 2019	Canonical Correlation Analysis	mNN pairing	
Liger	Welch et al. Cell 2019	Non-negative matrix factorization	KNN graph	
Conos	Barkas et al. Nat Methods 2019	Joint PCA	Inter/Intra- dataset edges	
scGen	Lotfollahi et al. Nat Methods 2019	Variational autoencoder	Decoder	Requires cell type annotation in both datasets
totalVI	Gayoso et al. bioRxiv 2019	Variational inference	Generative model	Requires multi-omic data from the same cells
BBKNN	Polanski et al. Bioinformatics 2019	PCA	Batch balanced graph construction	Bad alignment during testing
Cusanovich2018	Cusanovich et al. Cell 2018	PCA	mNN pairing	Code unavailable
gimVI	Lopez et al. arXiv 2019	Variational inference	Generative model	No implementation for right log-likelihood distribution

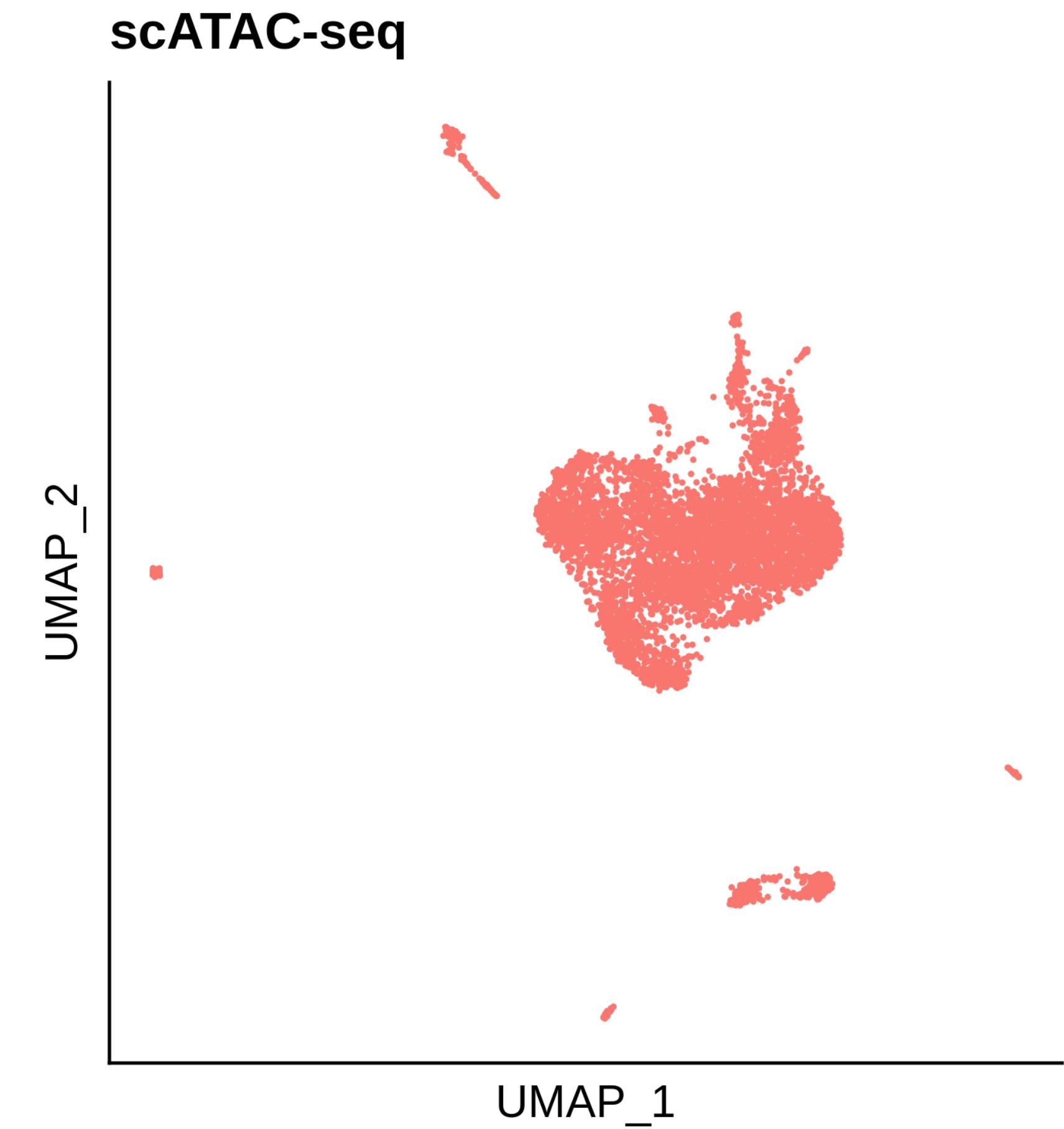
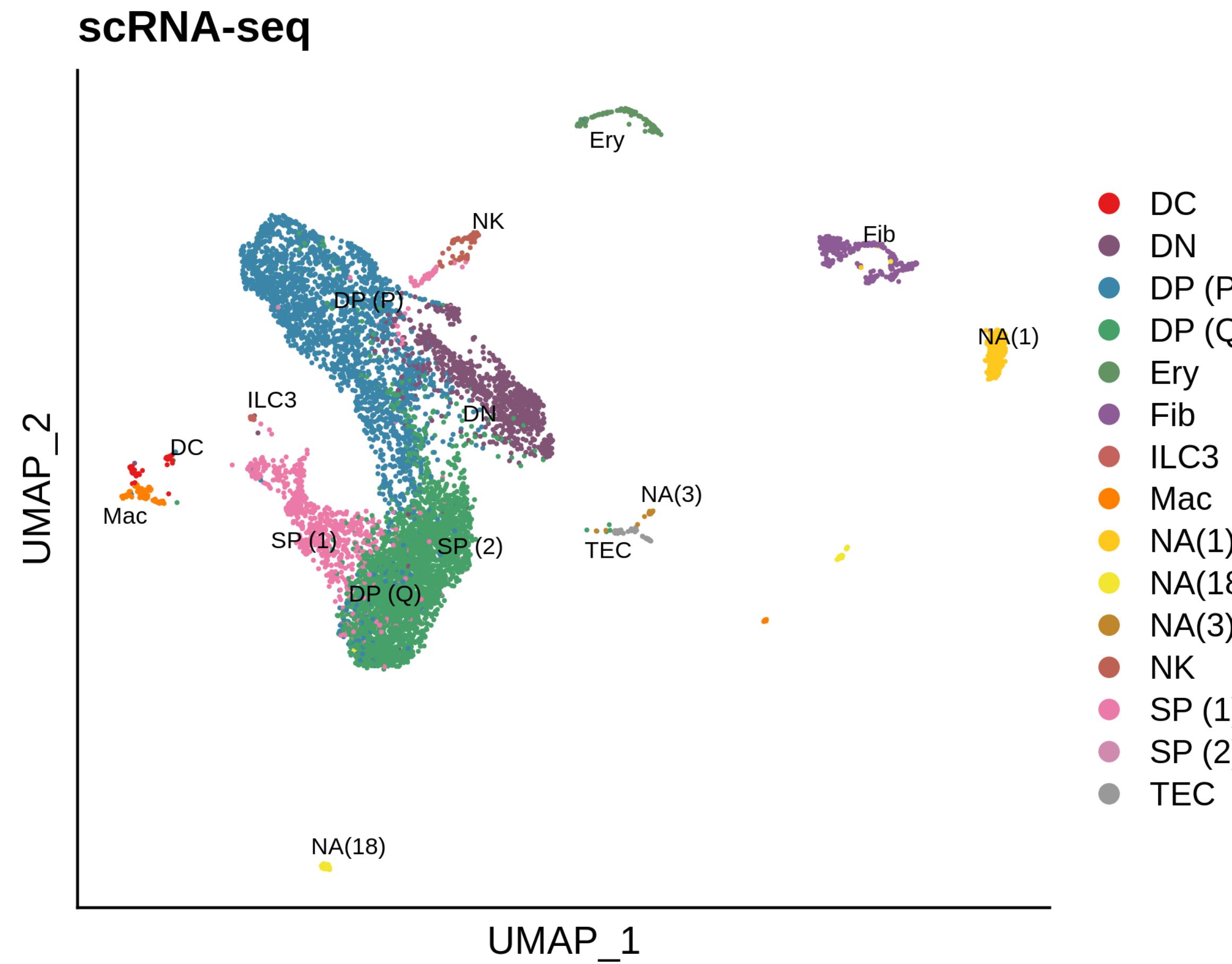
# Label transfer output - PBMC dataset



# Agreement with structure of ATAC-seq data

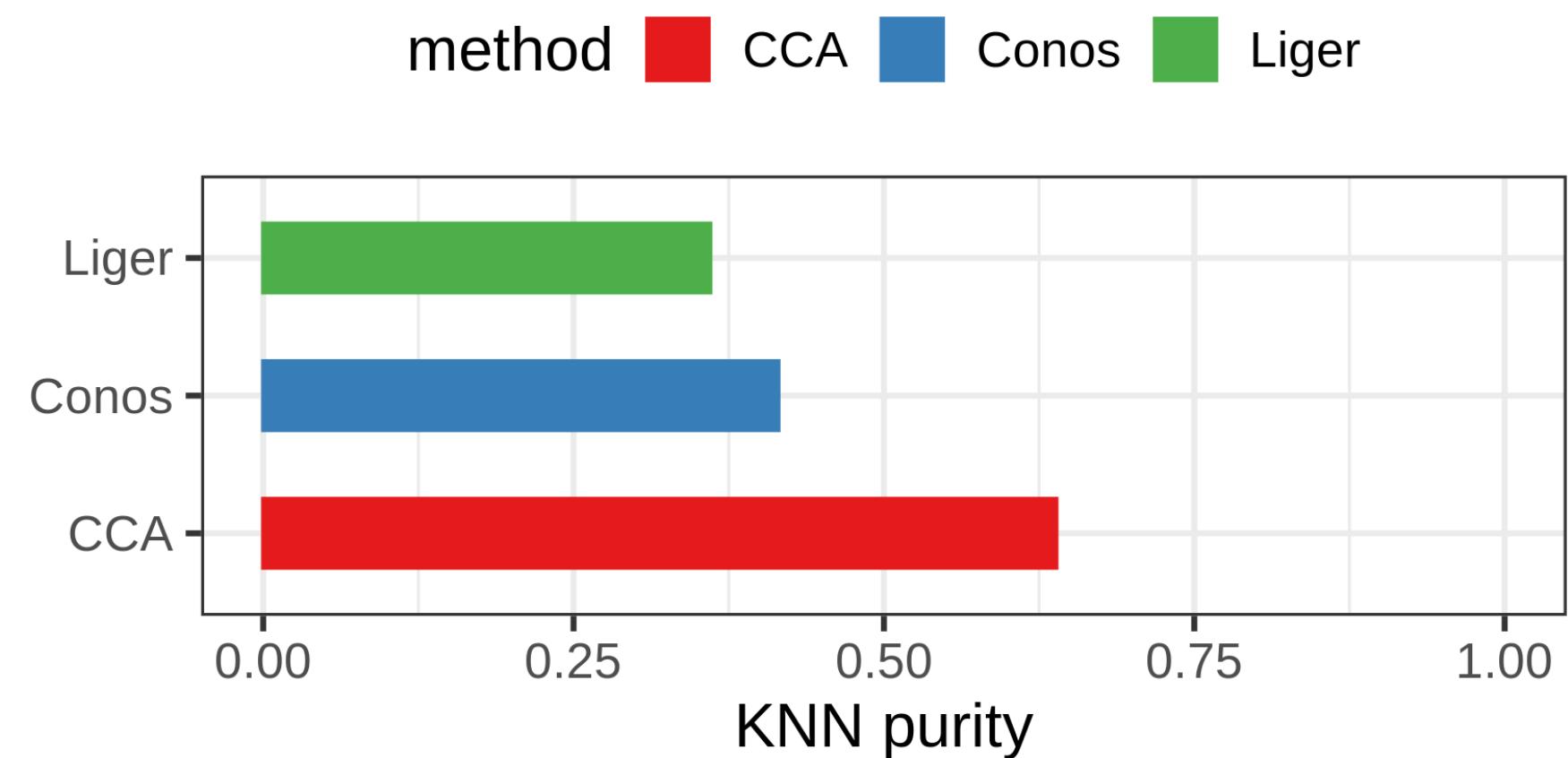
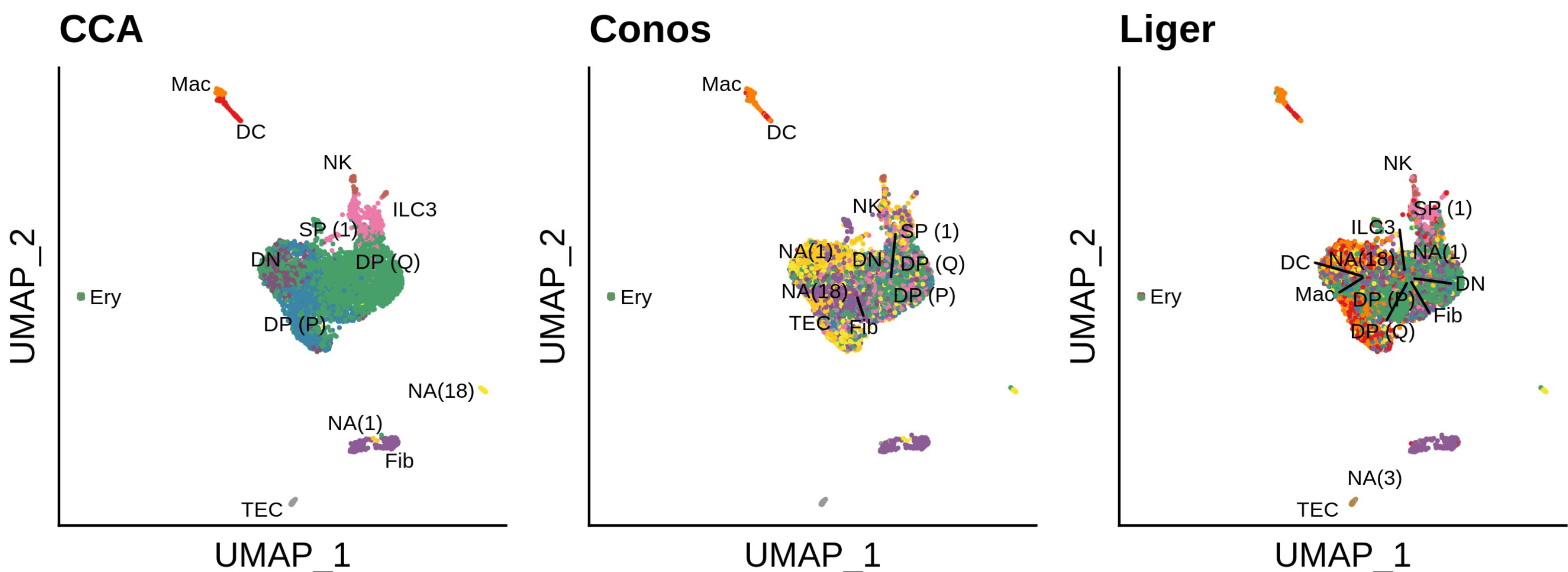


# Developing thymus dataset



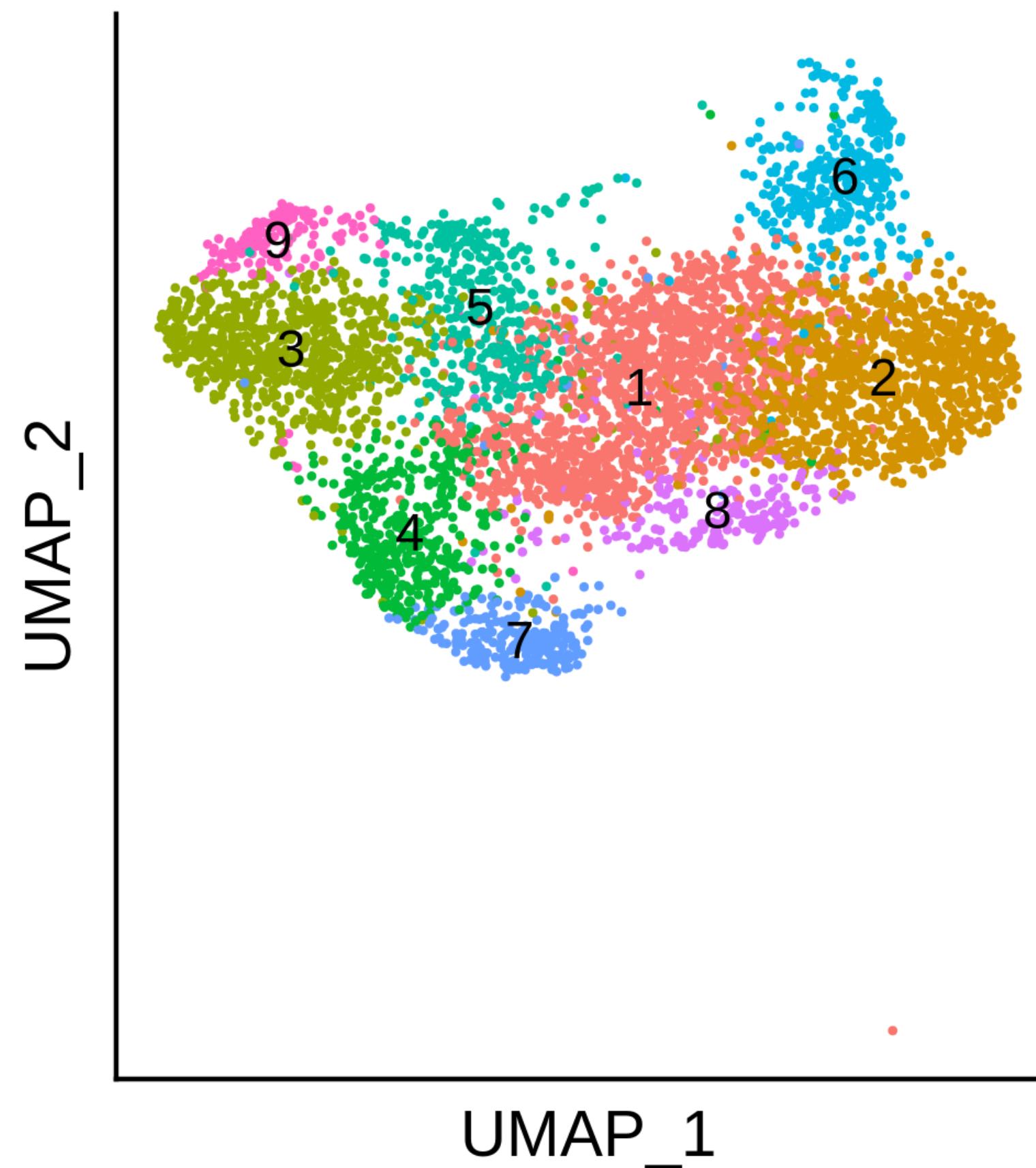
# Thymus - count GMAT

- DC
- DN
- DP (P)
- DP (Q)
- Ery
- Fib
- ILC3
- Mac
- NA(1)
- NA(18)
- NK
- SP (1)
- TEC

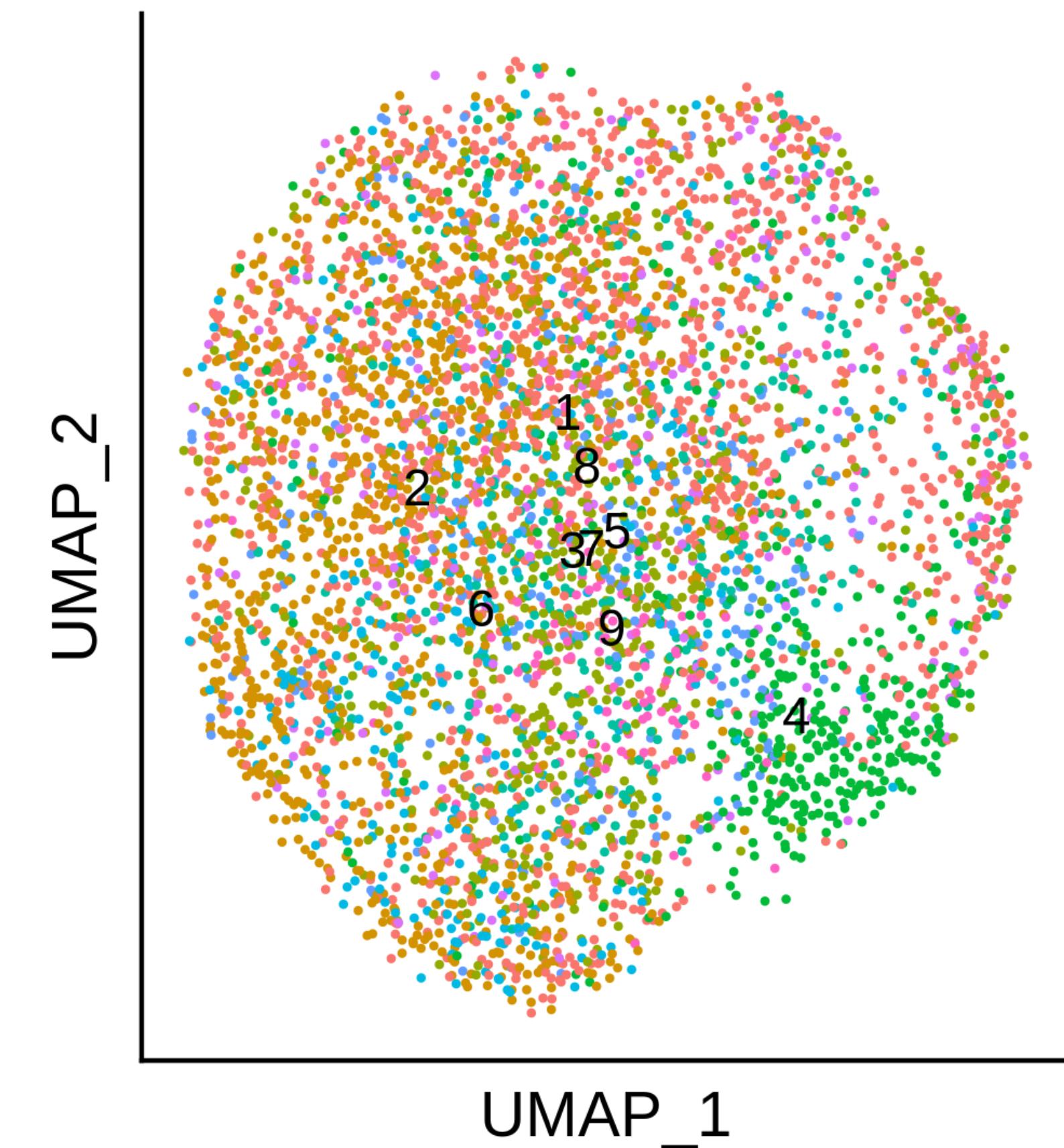


# Binarizing gene-level features maintains structure of full genome accessibility

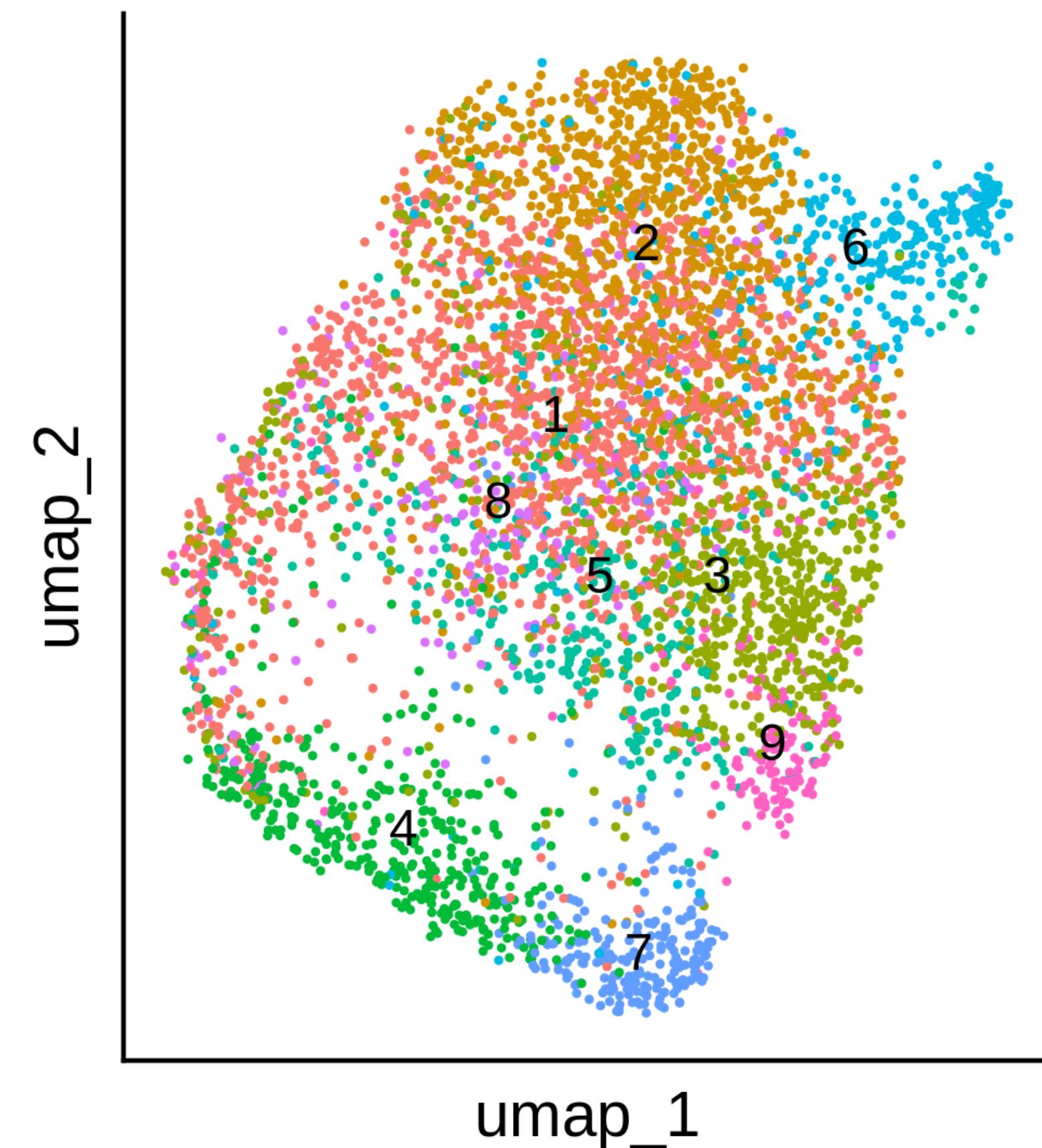
Genomic bins



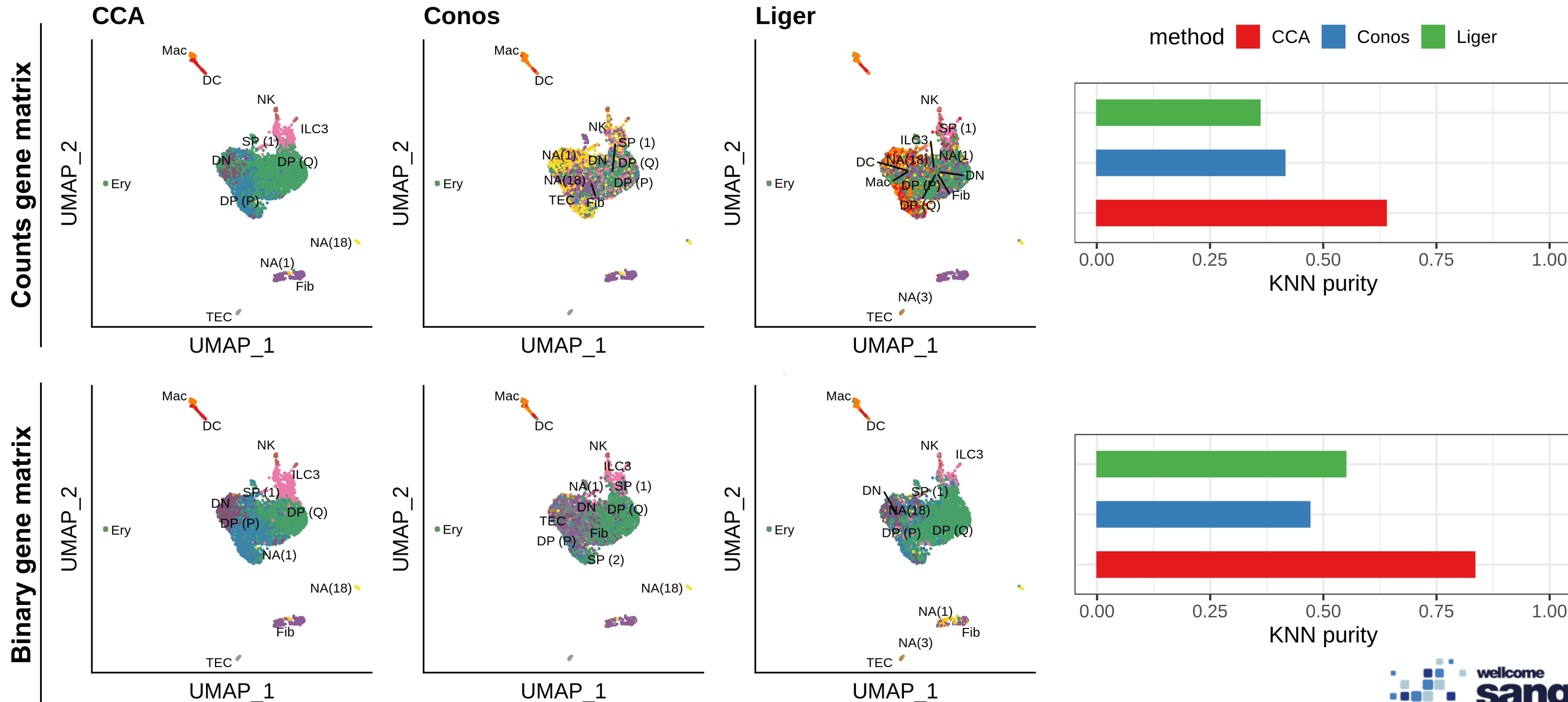
Gene accessibility (counts)



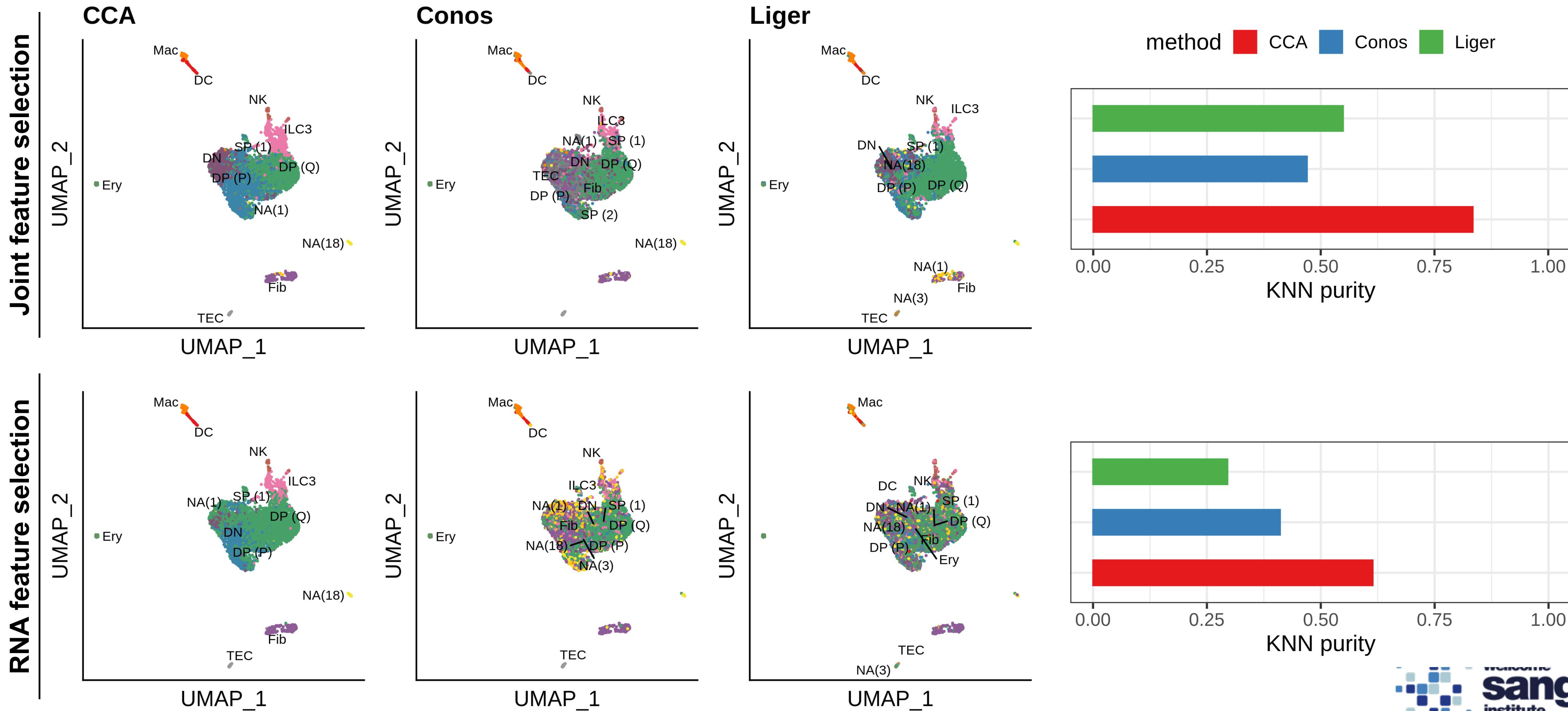
Gene accessibility (binary)



# Binarizing gene-level features increases quality of label transfer



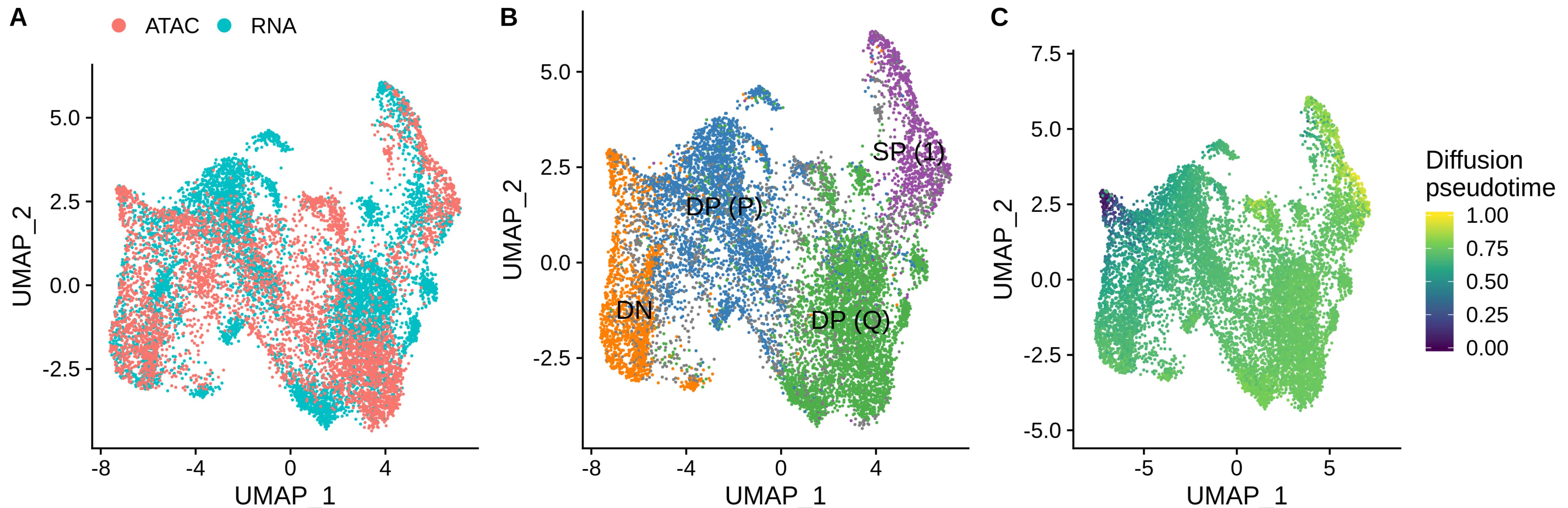
# Joint feature selection increases quality of label transfer



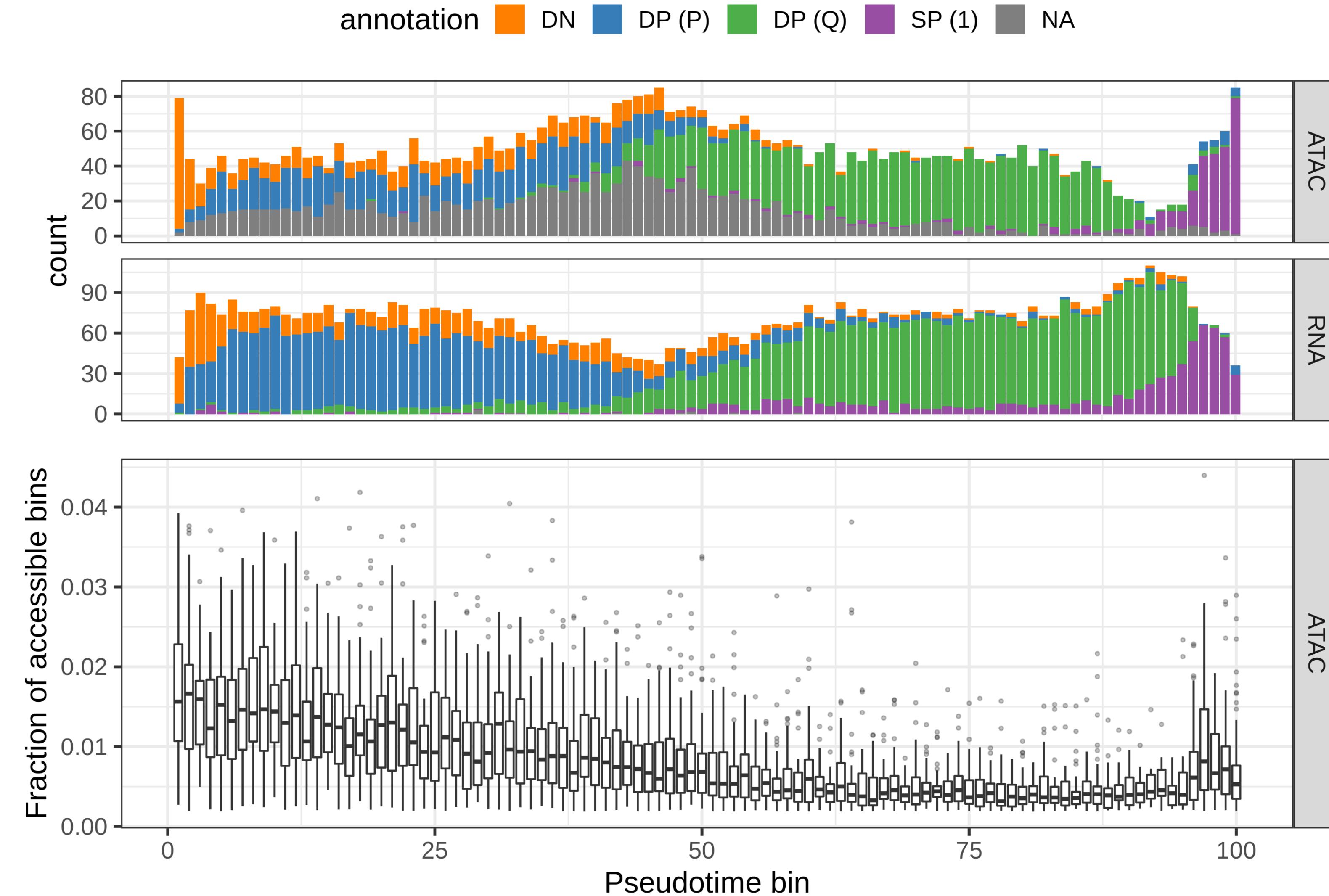
# Learnings

- CCA seems to outperform the other tested label transfer methods
- Binarization of the gene level accessibility data allows alignment of subclusters
- Joint feature selection is better

# Integration allows cell type prediction and trajectory inference on ATAC cells



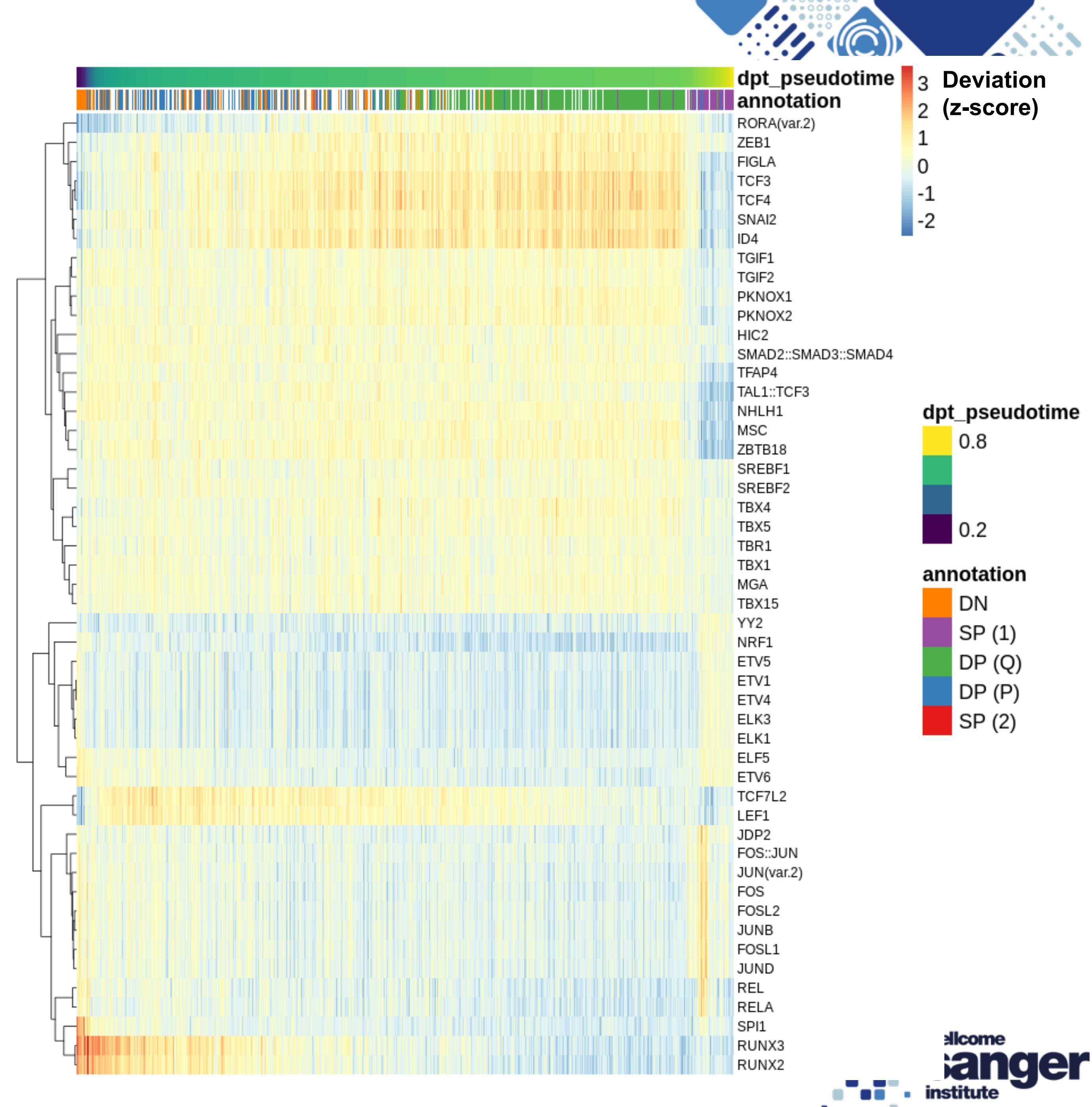
# Integration allows cell type prediction and trajectory inference on ATAC cells



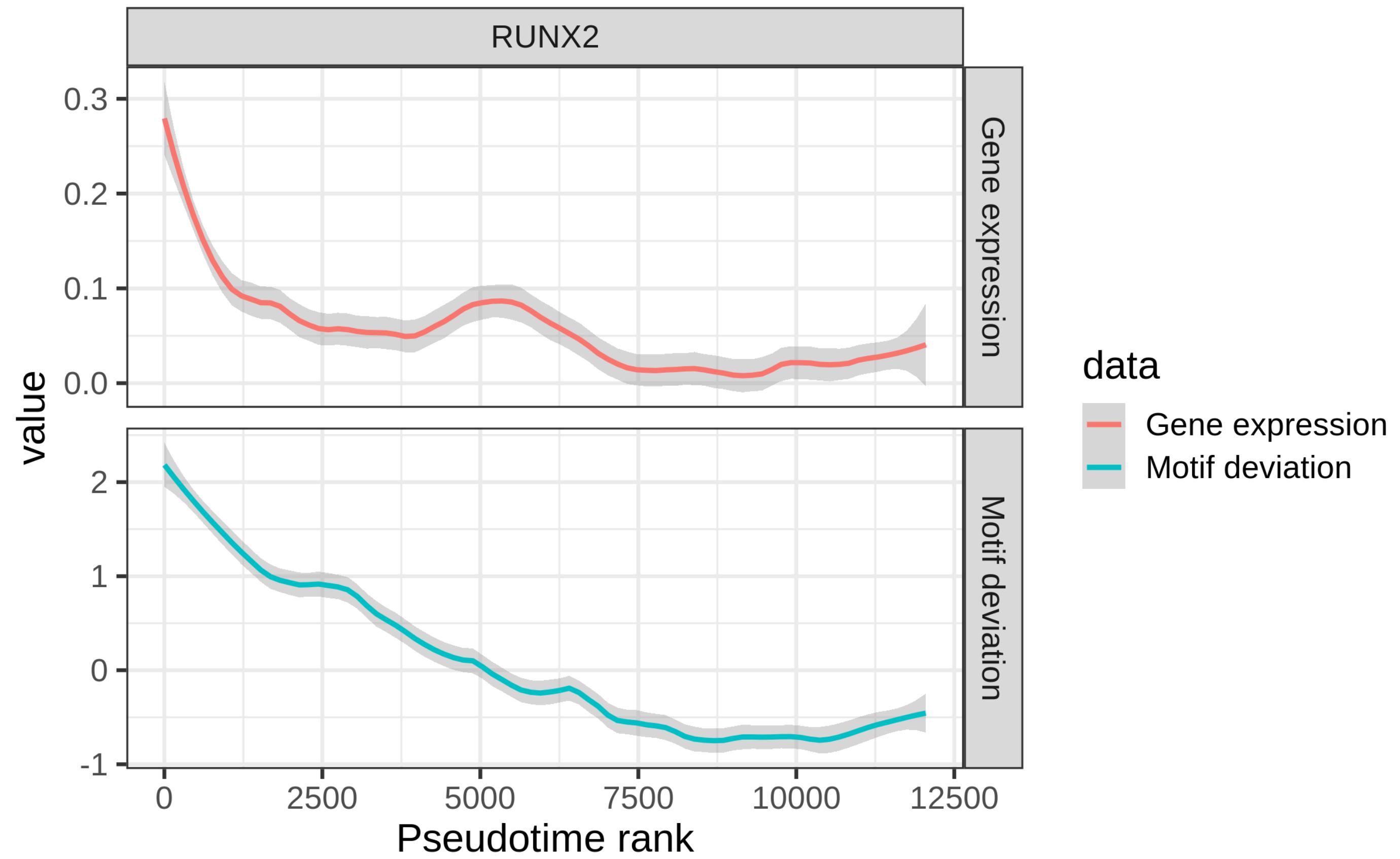
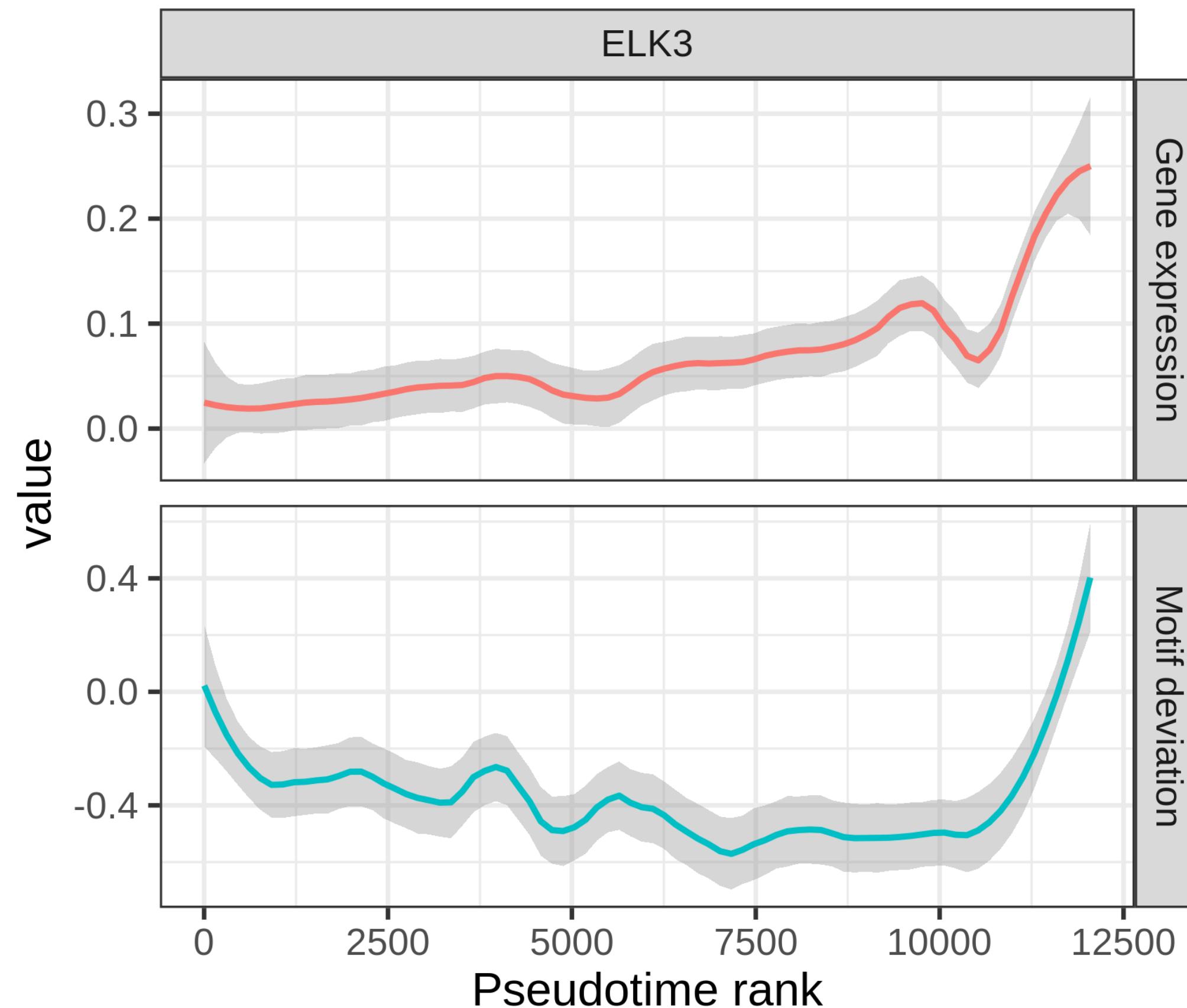
# Motif accessibility analysis

Using ChromVAR (Schep et al. 2017)

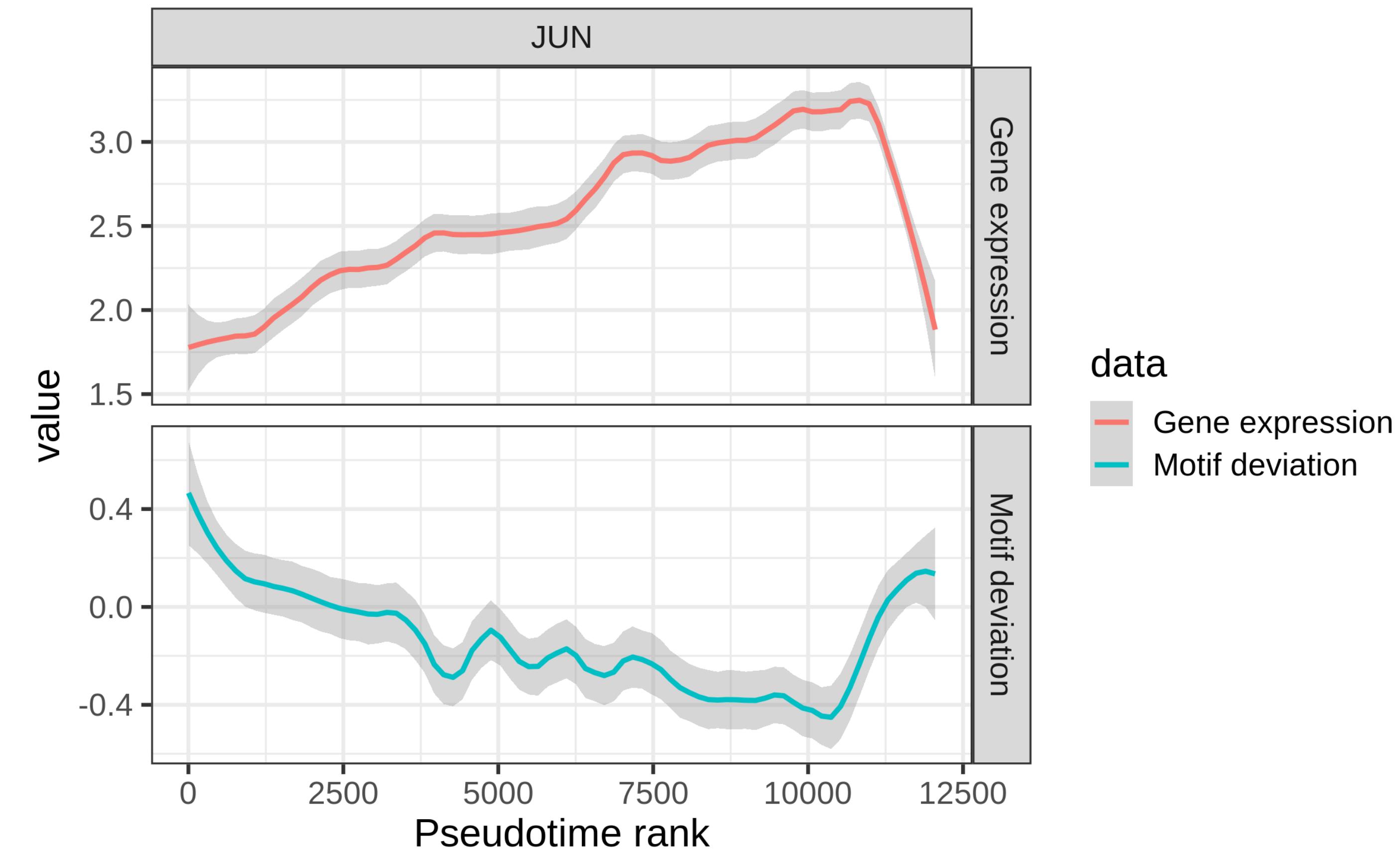
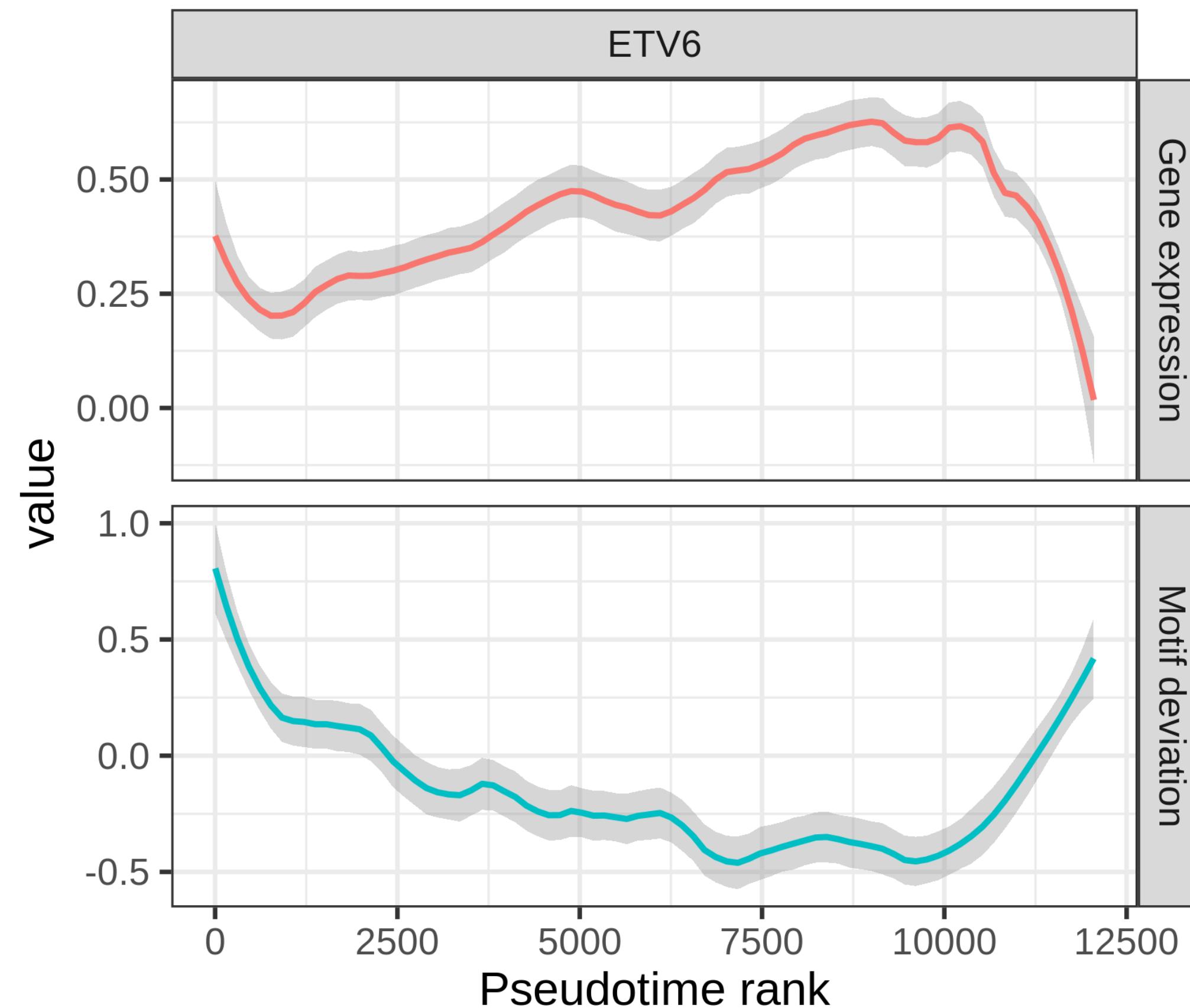
- Calculates enrichment/depletion of TF motif accessibility accounting for coverage and GC bias
- Identifies significantly variable TF motifs



# TF motif accessibility VS expression



# TF motif accessibility VS expression



# Acknowledgements

- Sarah
- Mirjana
- Cecilia
- JP
- Carlos
-