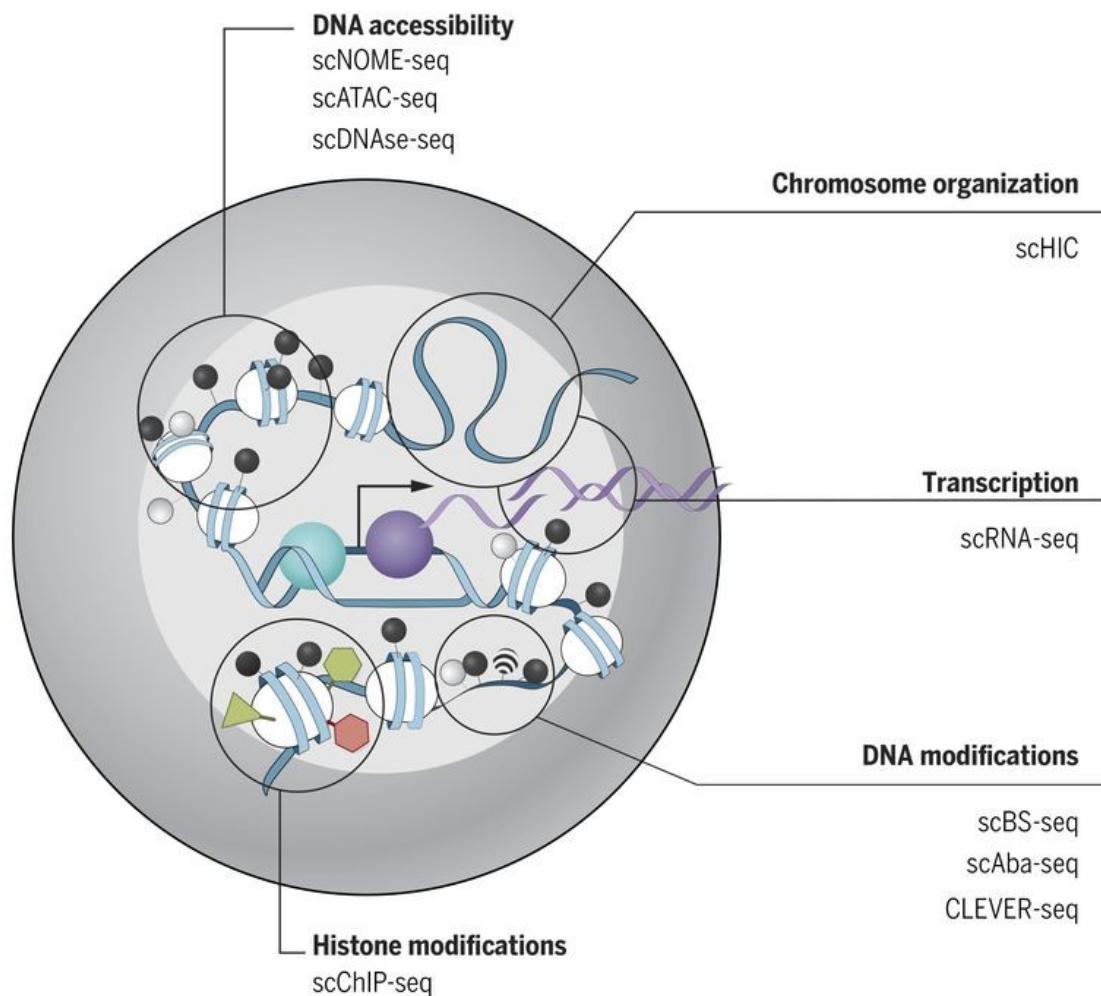
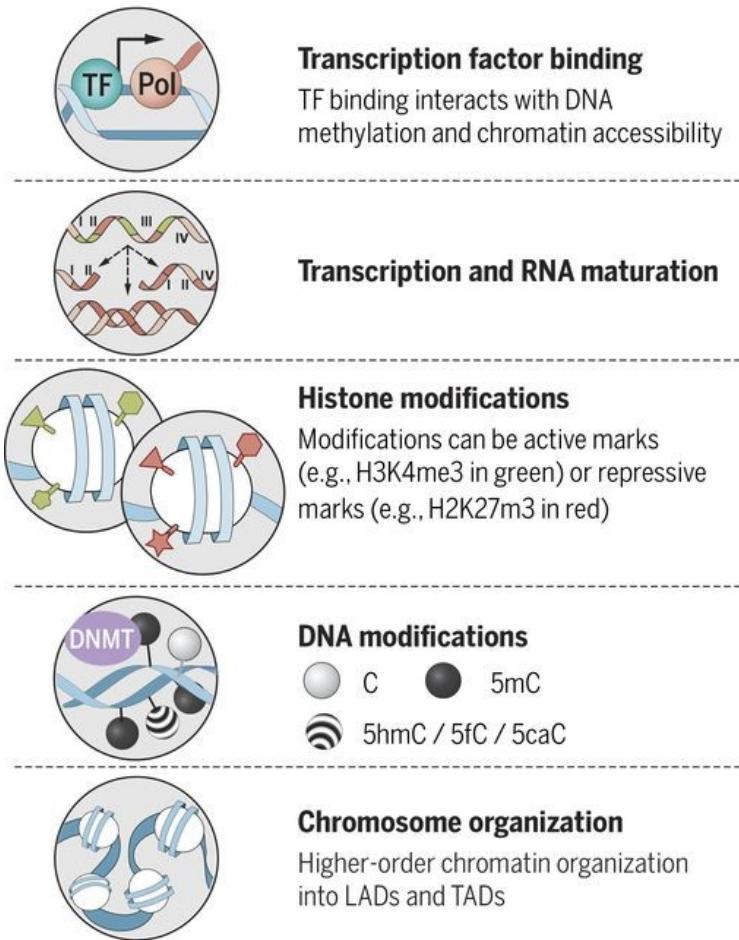
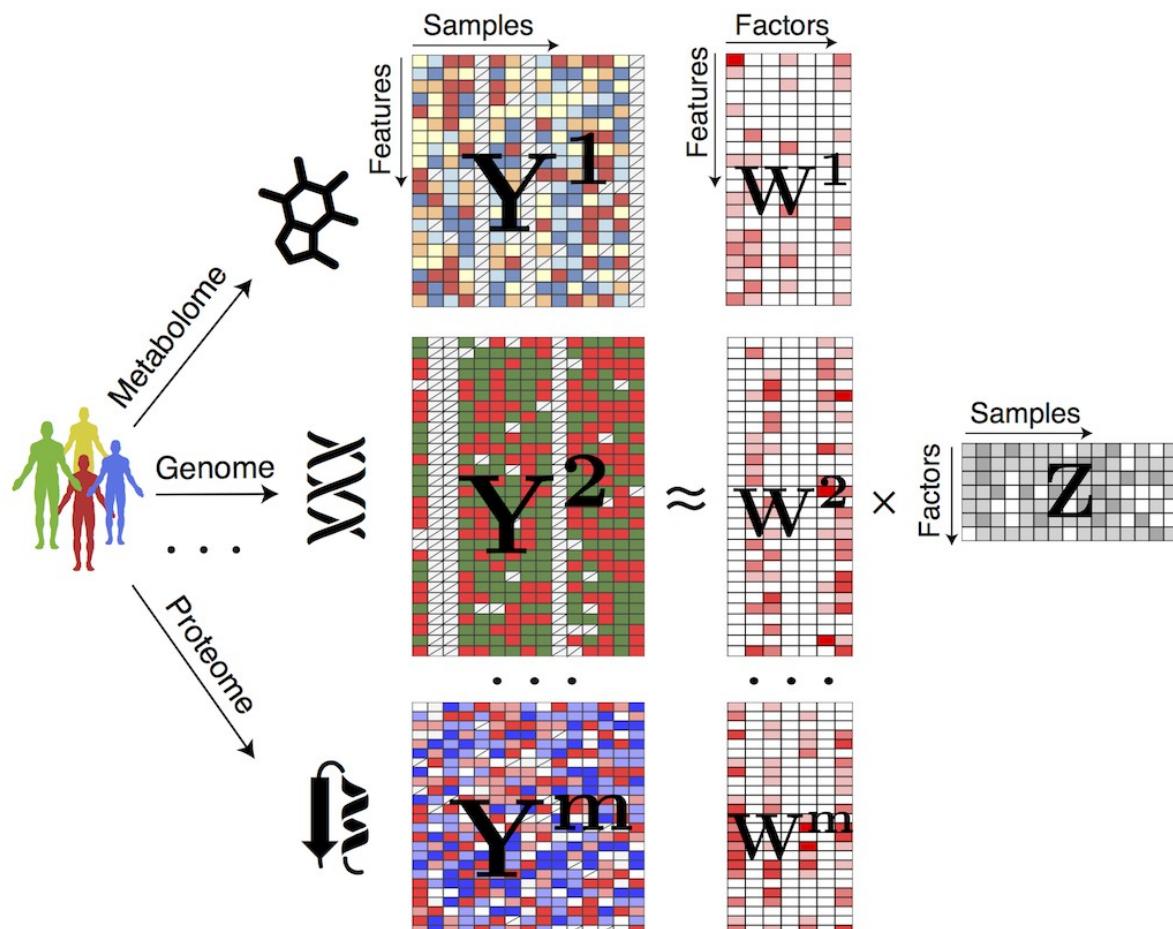
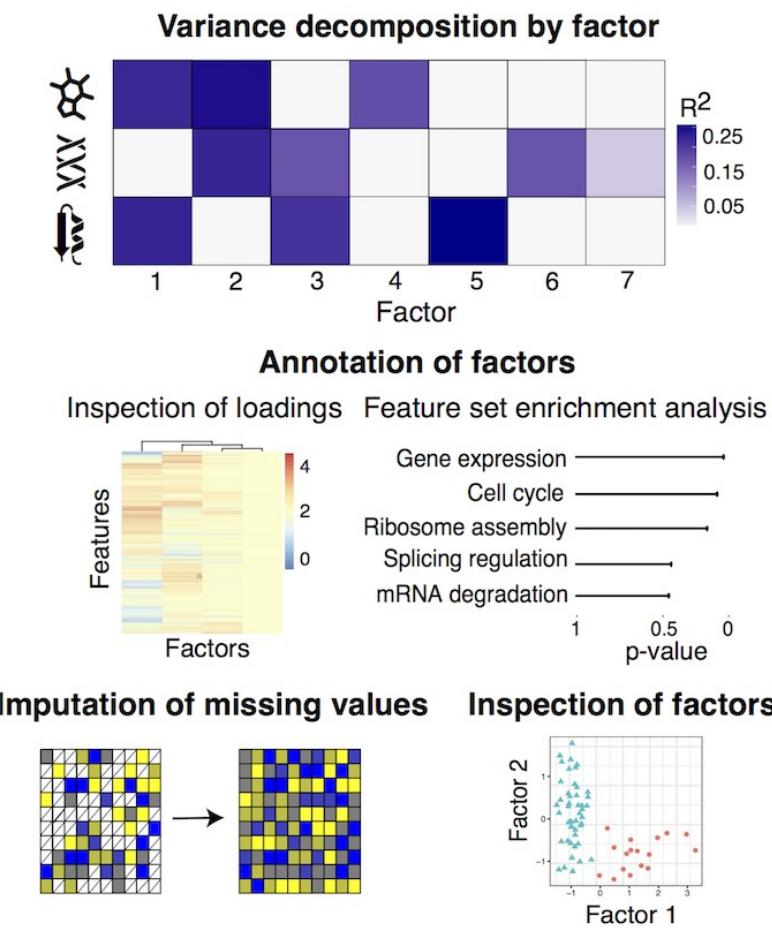


Unsupervised and Deep Learning Data Integration

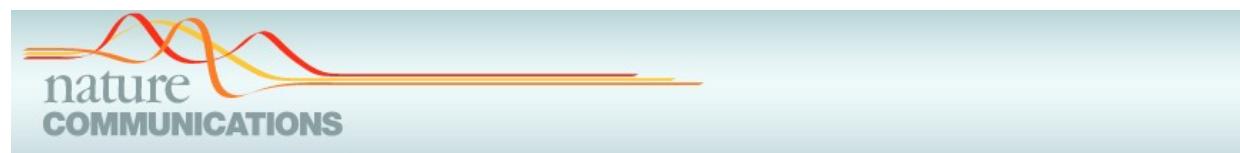
CZI Online Workshop
Nikolay Oskolkov, NBIS SciLifeLab
Lund, 08.04.2022



Step 1: train a MOFA model**Step 2: downstream analysis**

- Visualisation of samples / cells in a common across Omics factor space
- Annotation of factors using (gene set) enrichment analysis
- Imputation of missing values (as a result of data integration)
- Support of Omics with non-Gaussian distribution e.g. binary or count data

Argelaguet et al, 2018



NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-03149-4

ARTICLE

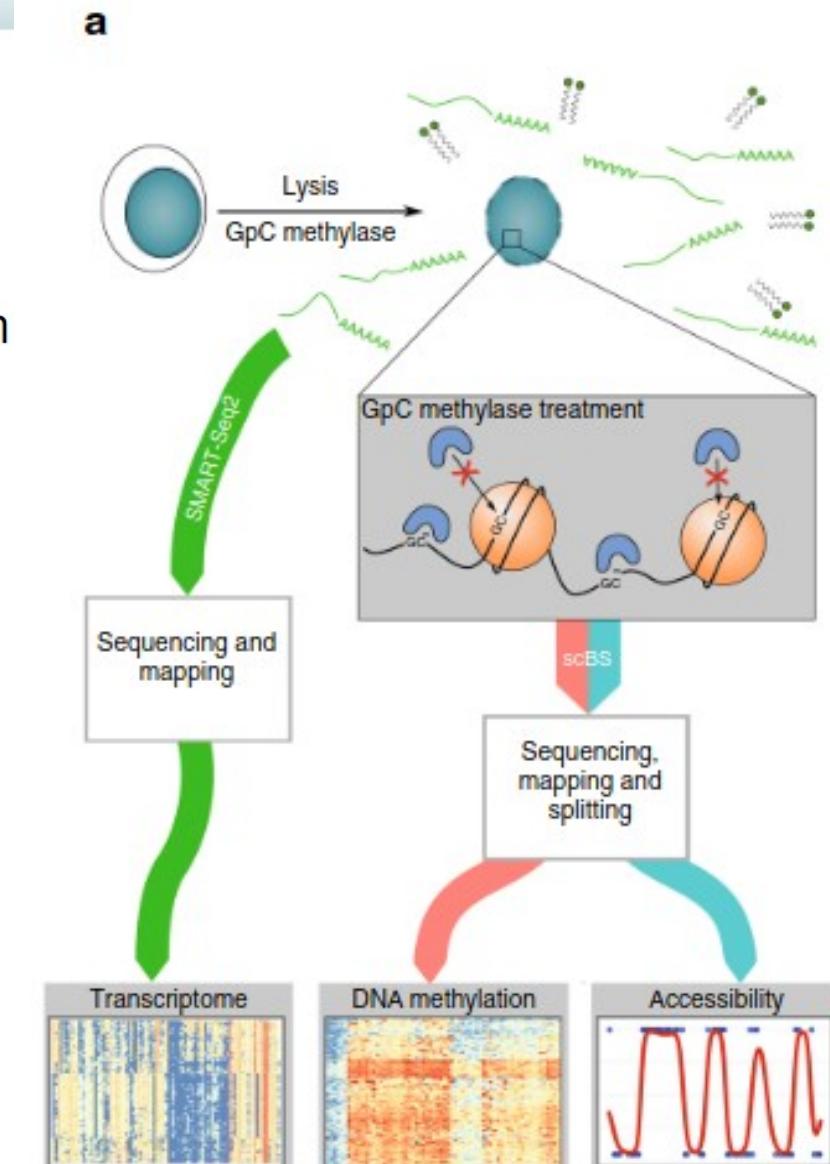
DOI: 10.1038/s41467-018-03149-4

OPEN

scNMT-seq enables joint profiling of chromatin accessibility, DNA methylation and transcription in single cells

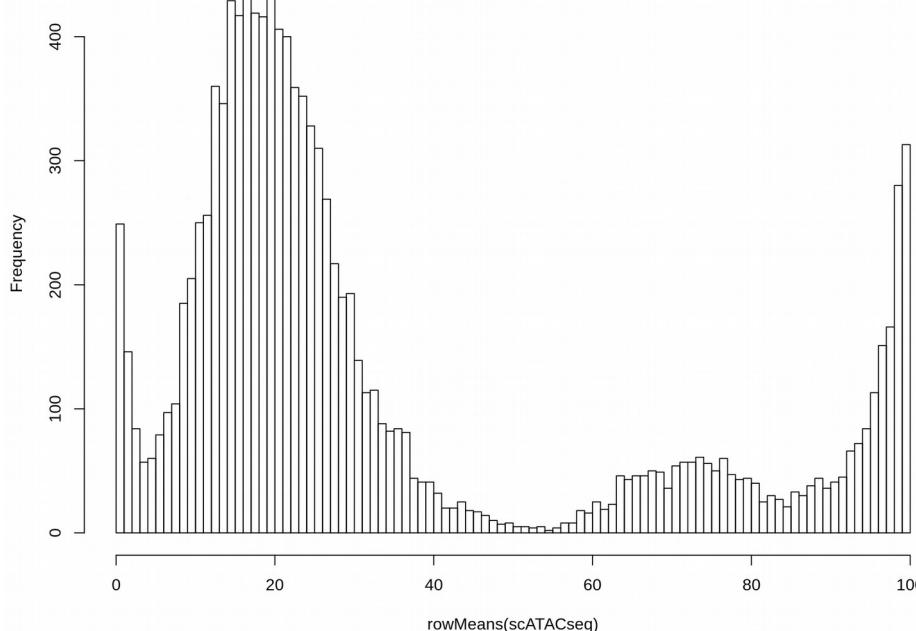
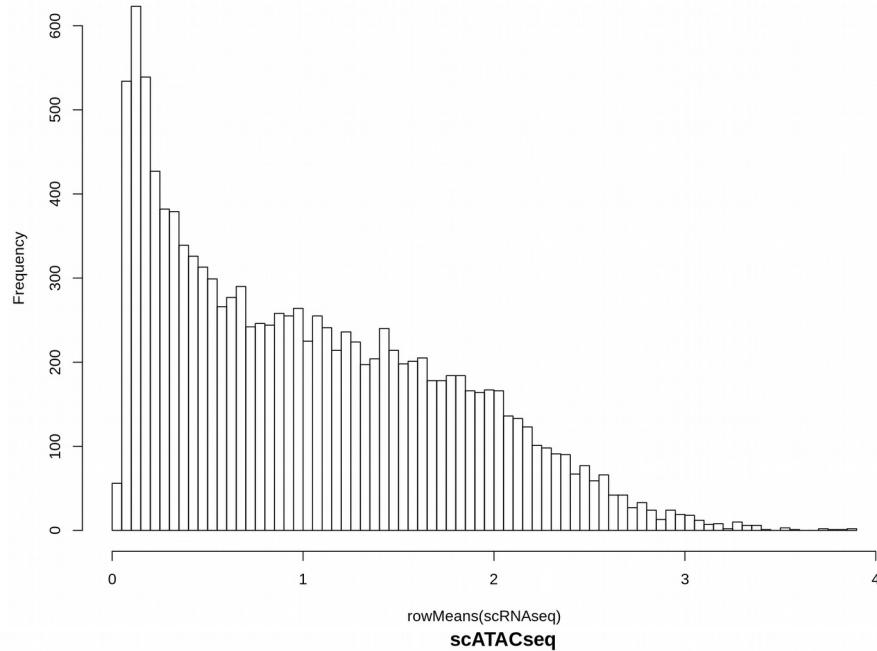
Stephen J. Clark¹, Ricard Argelaguet^{2,3}, Chantriolnt-Andreas Kapourani¹, Thomas M. Stubbs¹, Heather J. Lee^{1,5,6}, Celia Alda-Catalinas¹, Felix Krueger¹, Guido Sanguinetti⁴, Gavin Kelsey^{1,8}, John C. Marioni^{1,2,3,5}, Oliver Stegle¹, Wolf Reik^{1,5,8}

Parallel single-cell sequencing protocols represent powerful methods for investigating regulatory relationships, including epigenome-transcriptome interactions. Here, we report a single-cell method for parallel chromatin accessibility, DNA methylation and transcriptome profiling. scNMT-seq (single-cell nucleosome, methylation and transcription sequencing) uses a GpC methyltransferase to label open chromatin followed by bisulfite and RNA sequencing. We validate scNMT-seq by applying it to differentiating mouse embryonic stem cells, finding links between all three molecular layers and revealing dynamic coupling between epigenomic layers during differentiation.

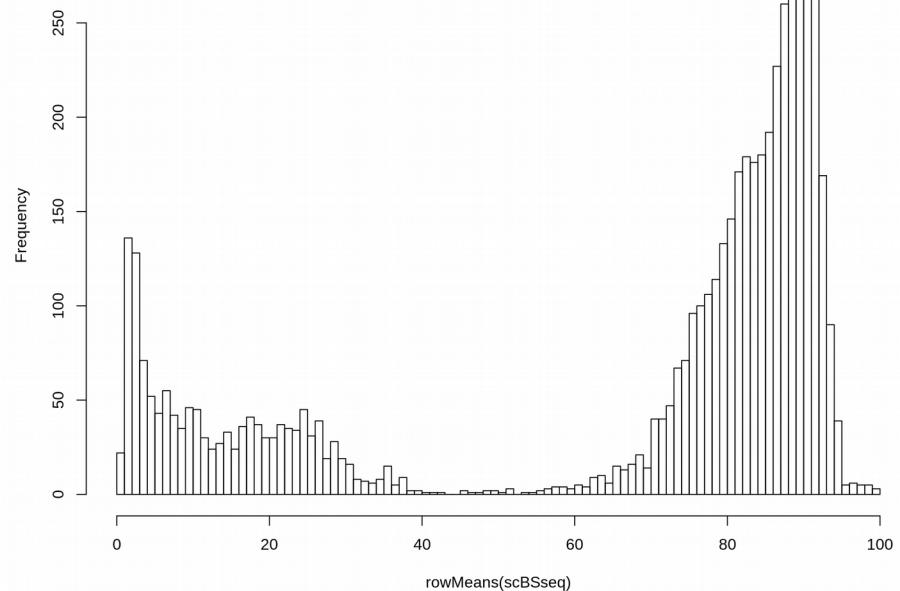


Mouse embryonic stem cells (ESC) +
embryoid bodies (EB) cells + control cells

scRNAseq



scBSseq



LIKELIHOOD

The probability of "B" being True, given "A" is True

PRIOR

The probability "A" being True. This is the knowledge.

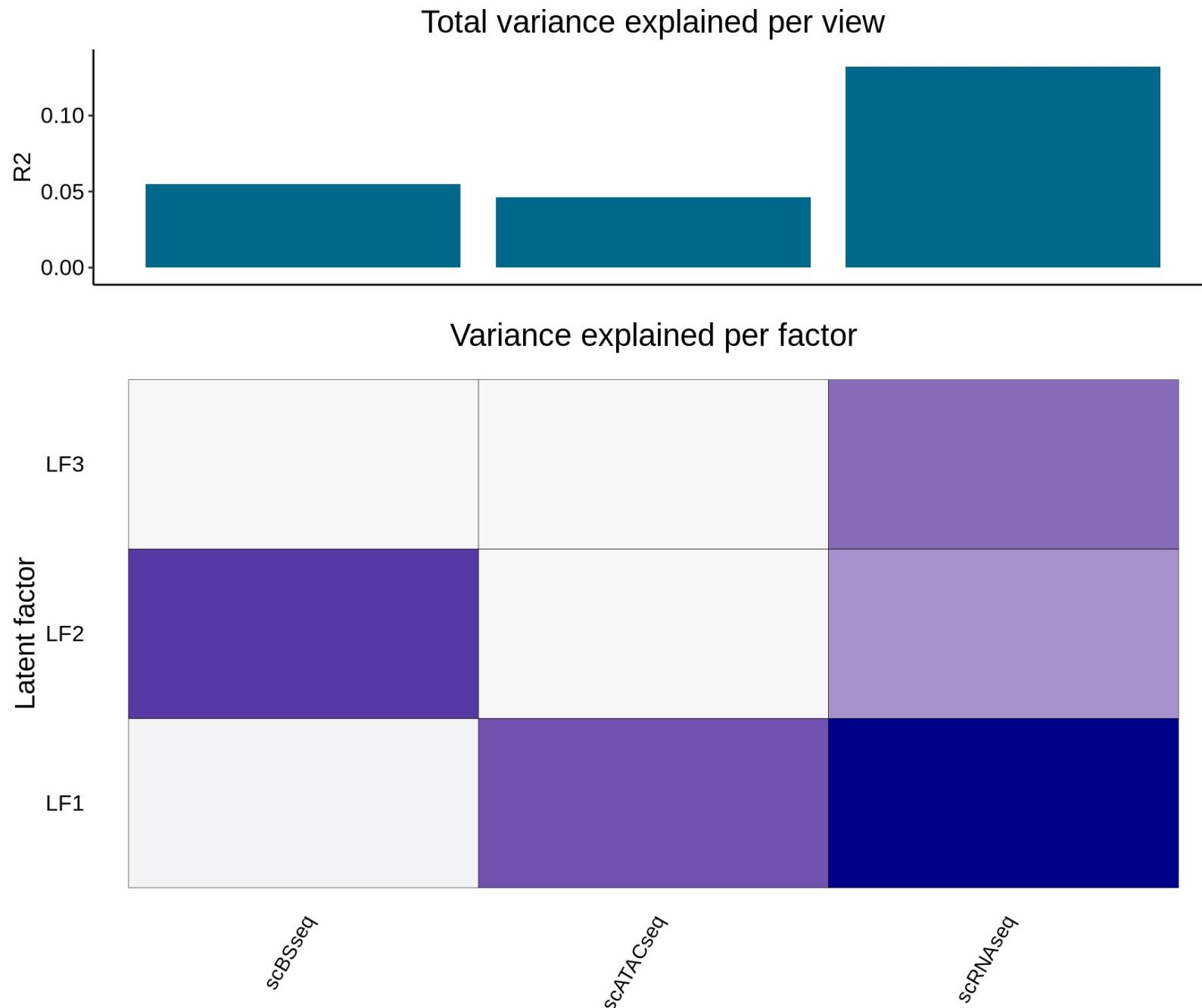
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

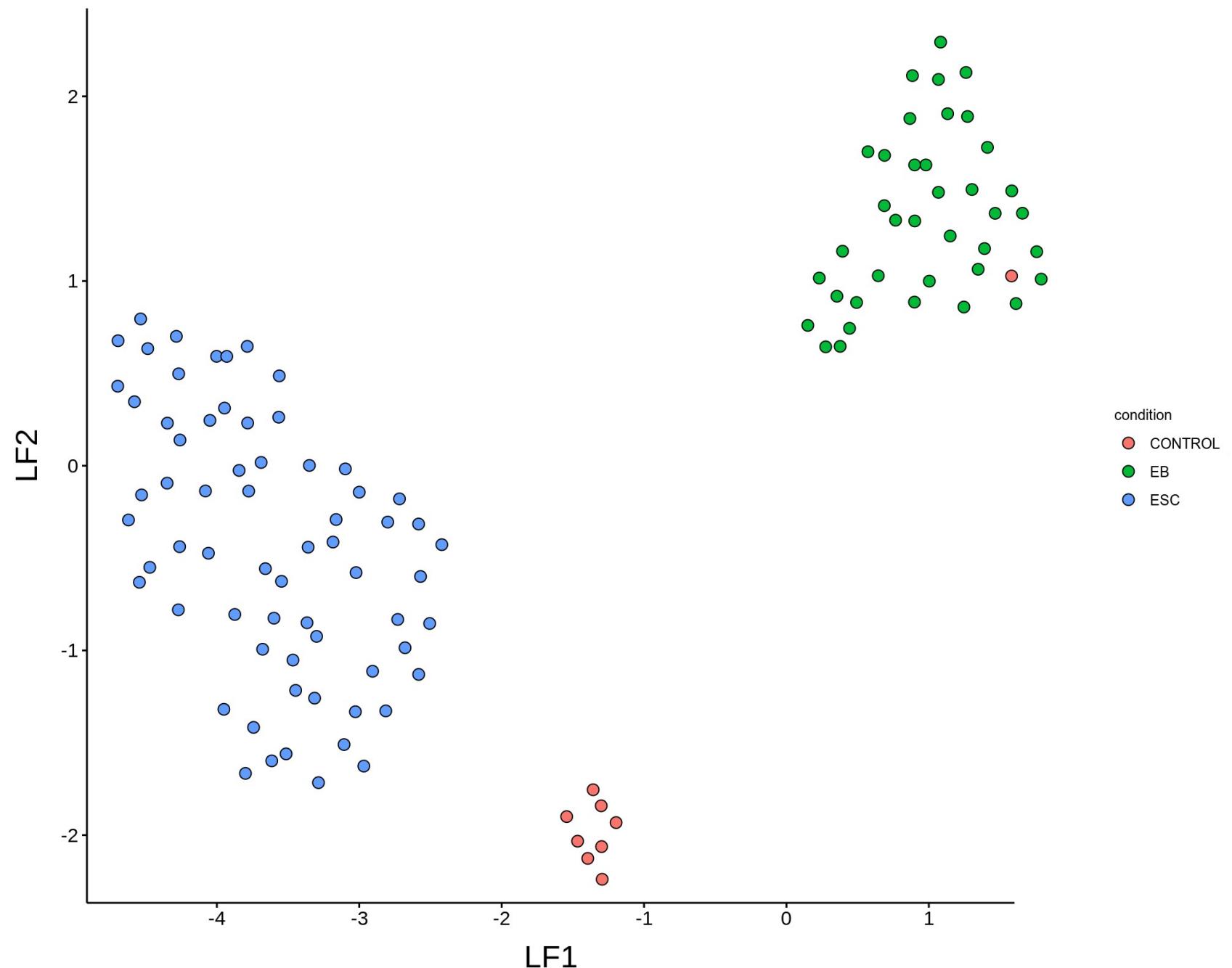
POSTERIOR

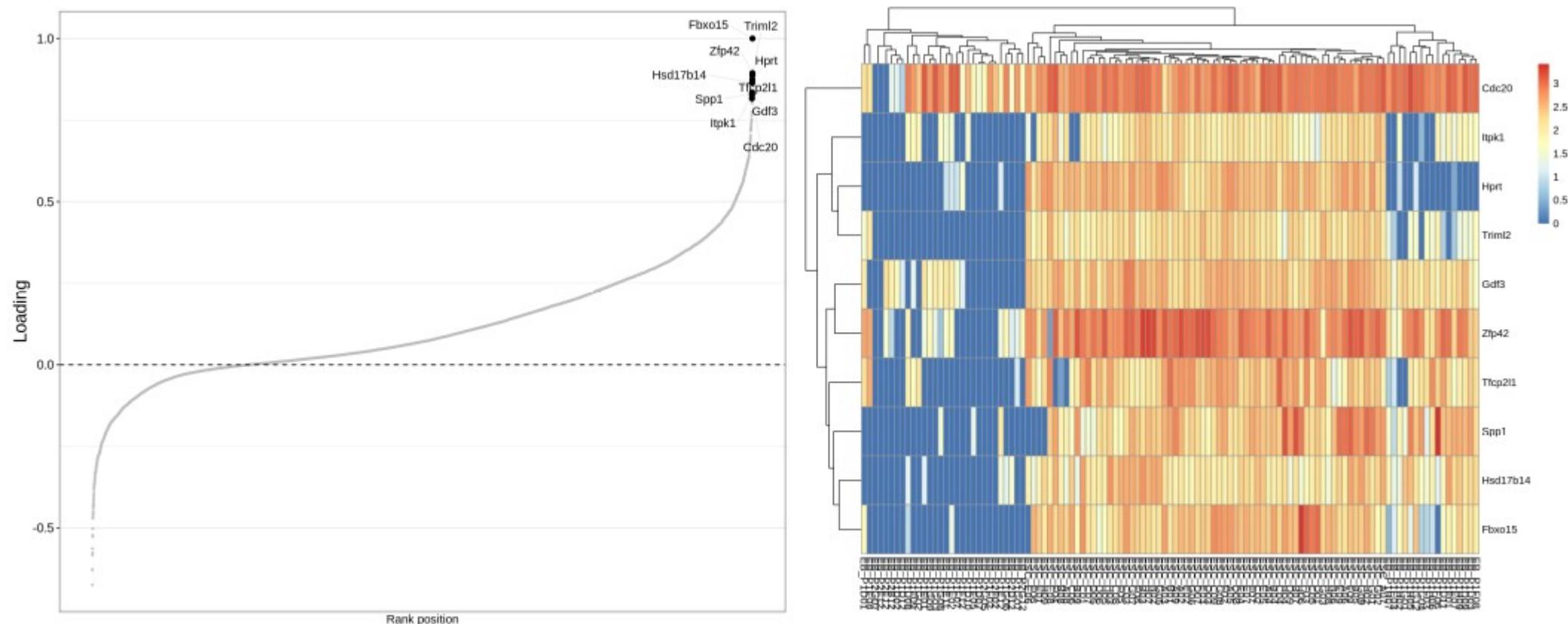
The probability of "A" being True, given "B" is True

MARGINALIZATION

The probability "B" being True.

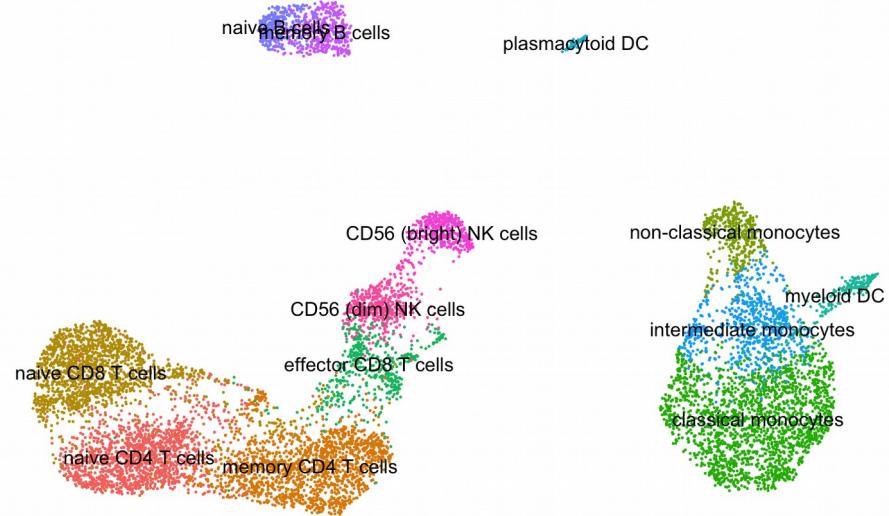




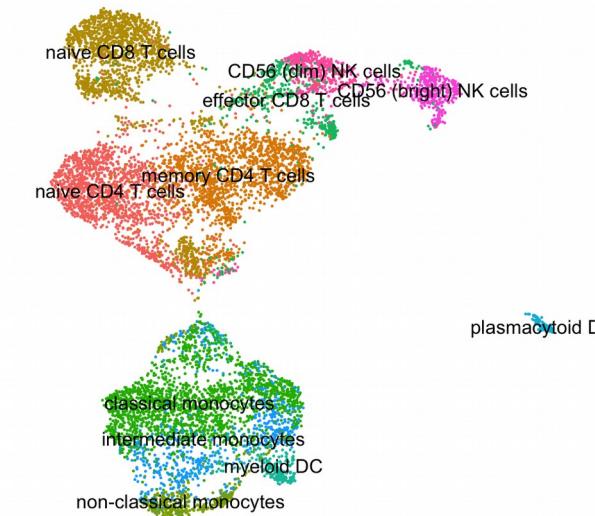


ESC and EB cells are separable on the heatmap built on loadings of the MOFA latent factors

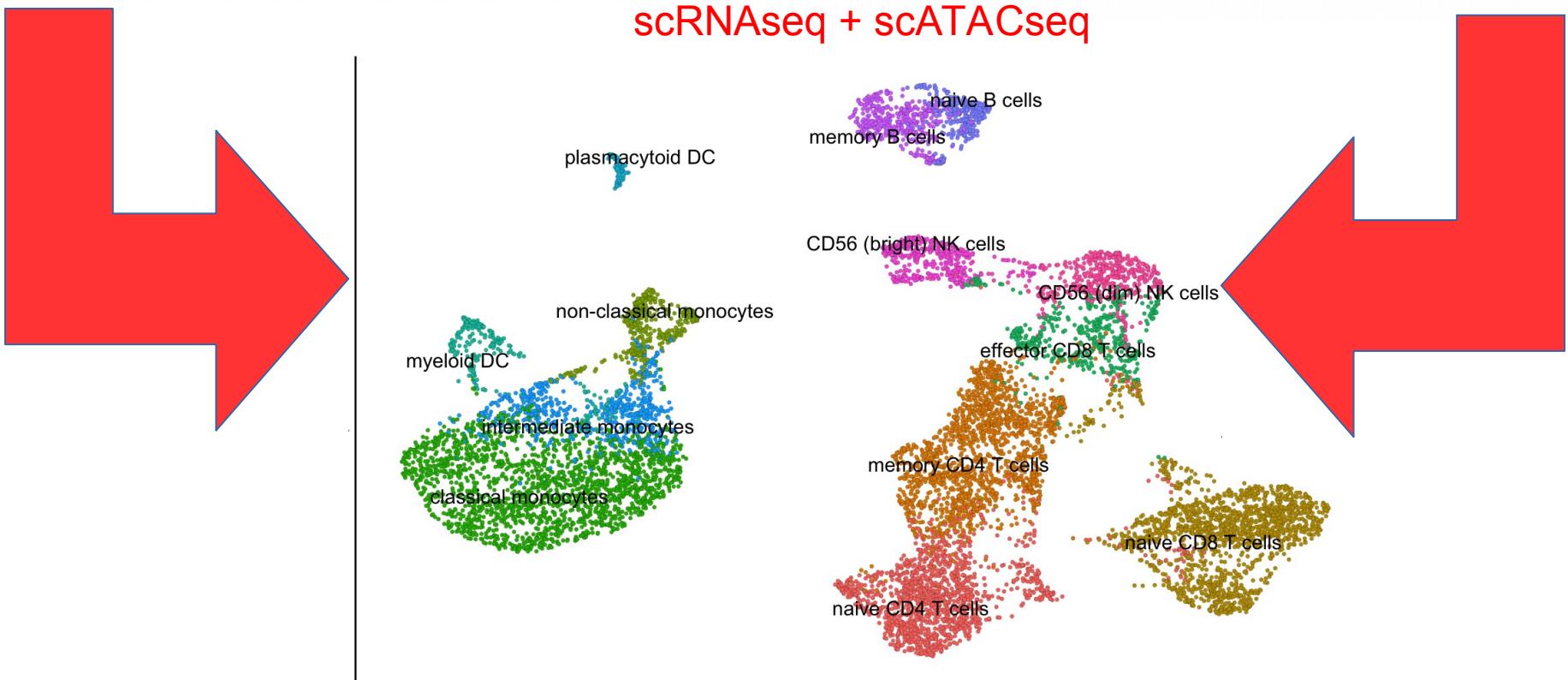
scRNAseq

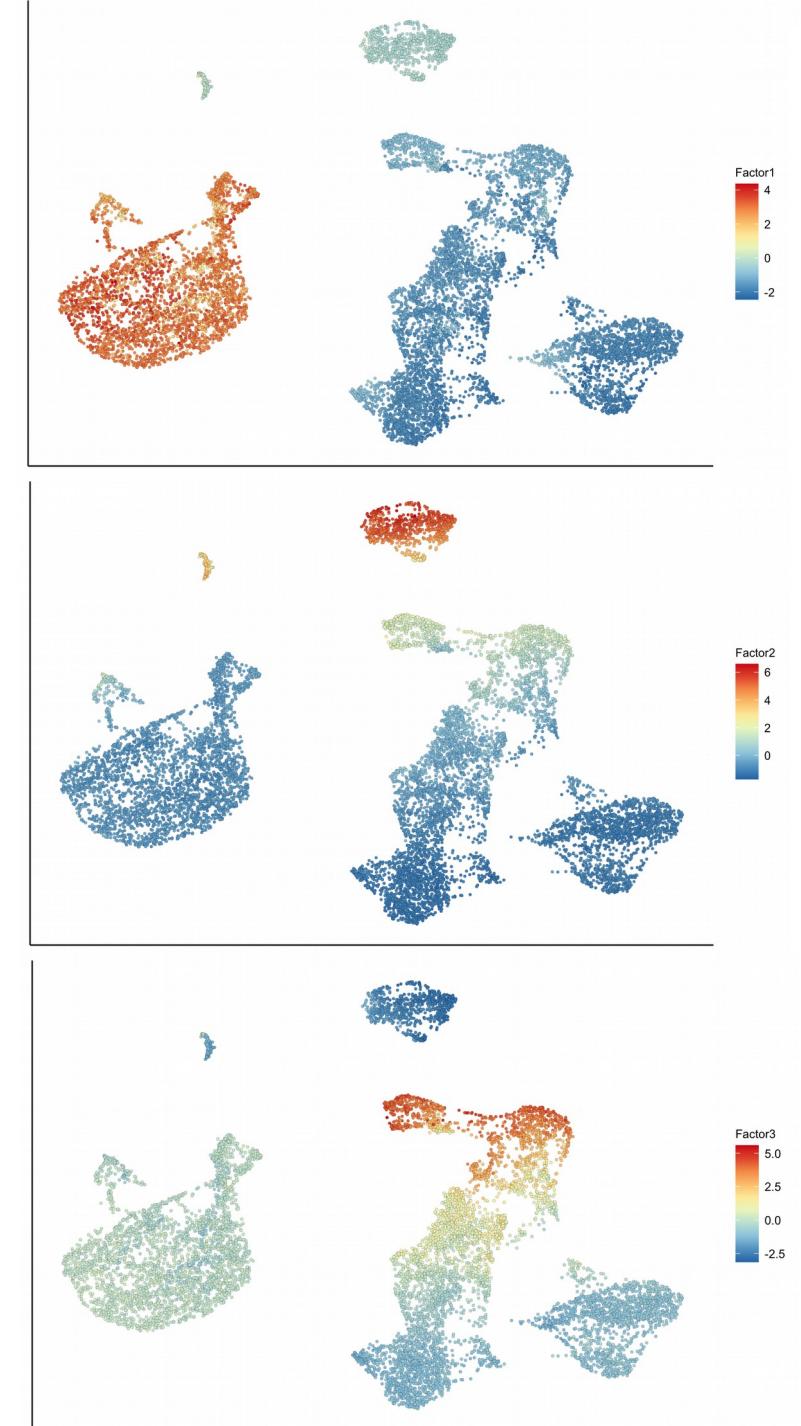
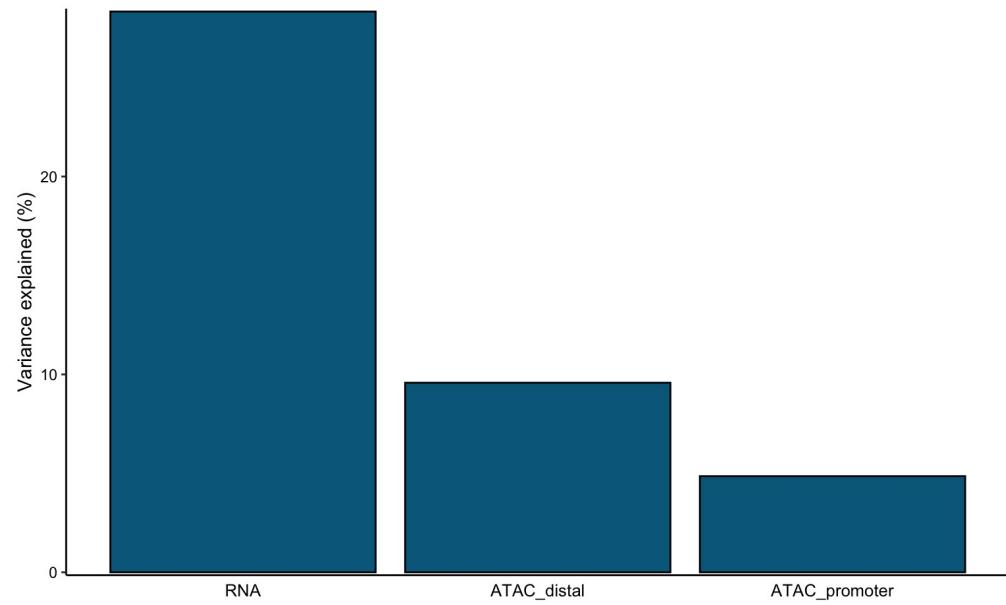


scATACseq



scRNAseq + scATACseq





Other Unsupervised Integrative OMICs Methods

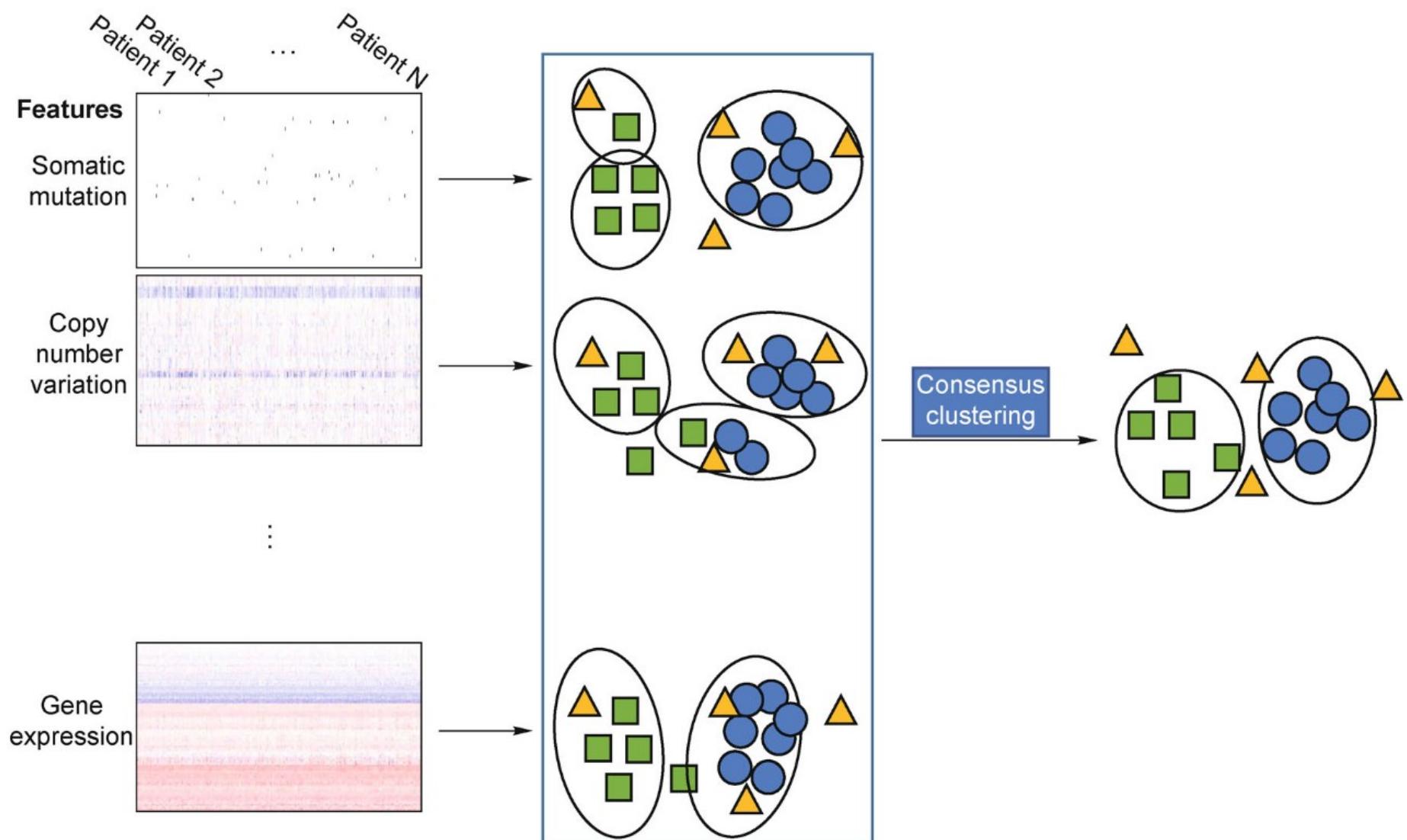
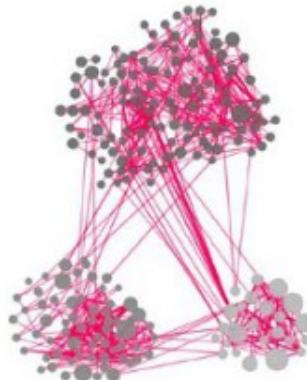
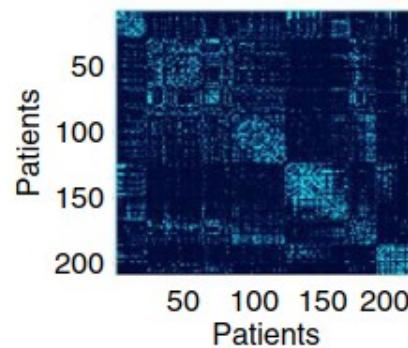
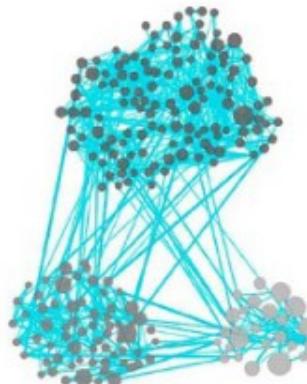
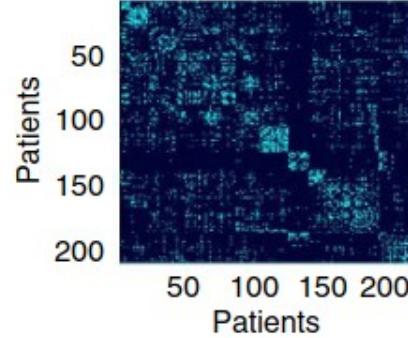
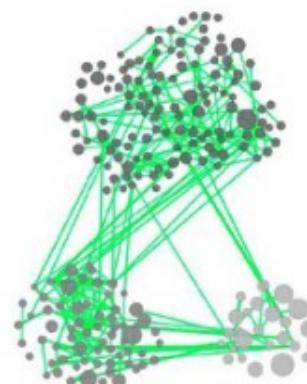
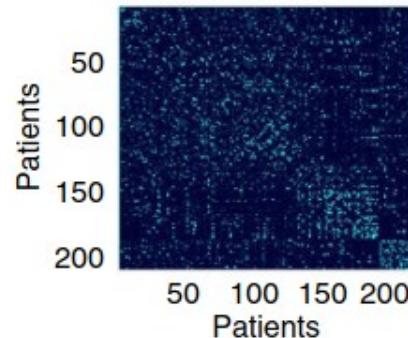
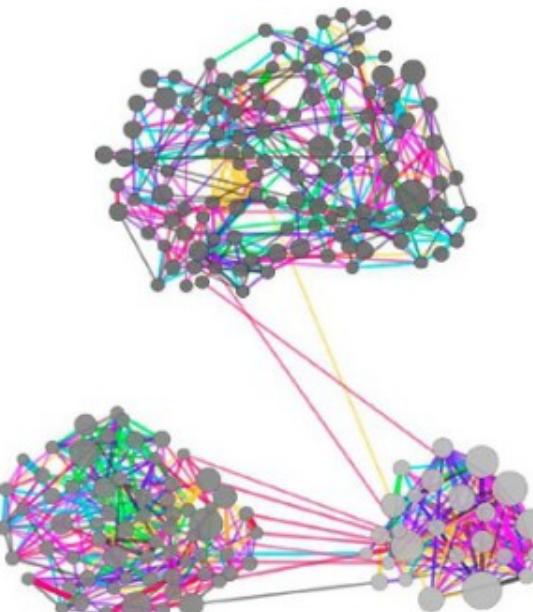
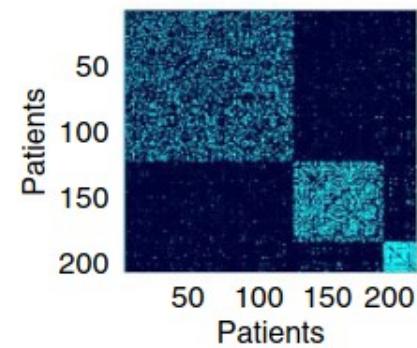


Figure 2. Clustering of clusters. This kind of methods first clusters in every single omics dataset and then integrates the primary clustering results into final cluster assignments.

a**b****c****d**

Patient subtype

1 2 3

Survival (months)

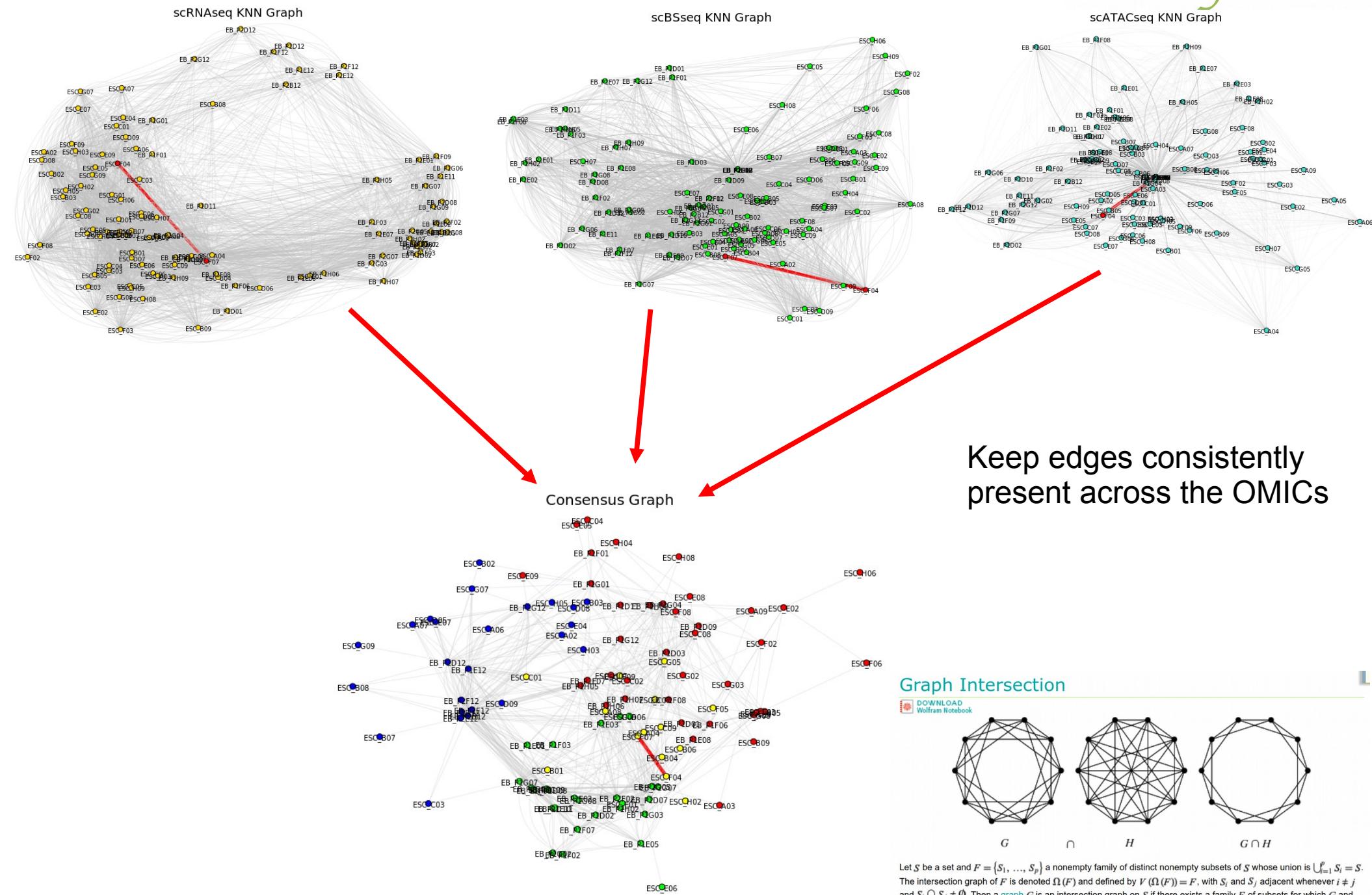
1 48 115

Similarly type

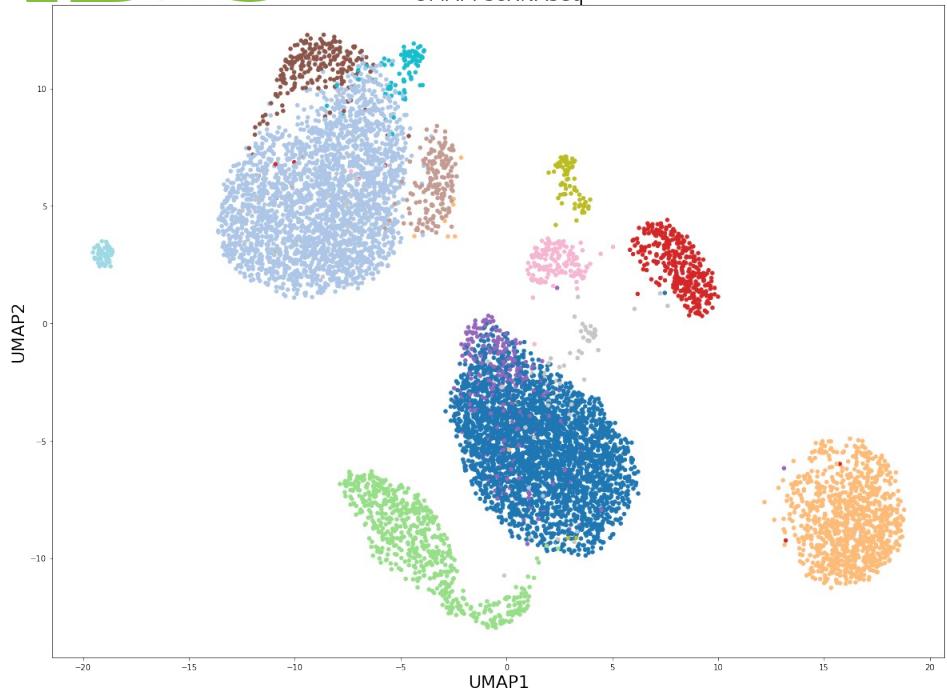
miRNA

DNA
methylation

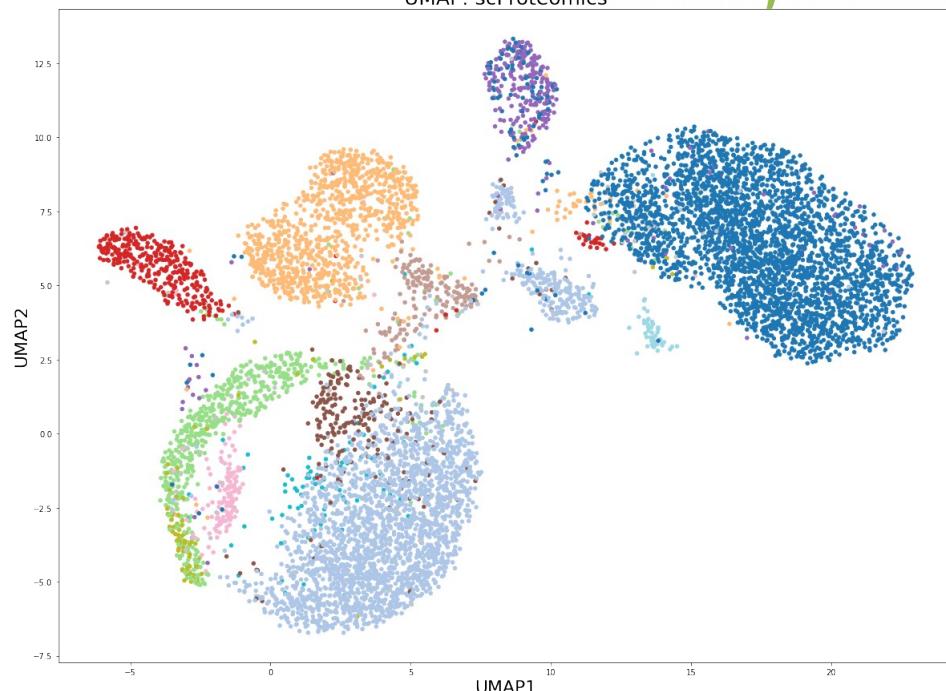
mRNA



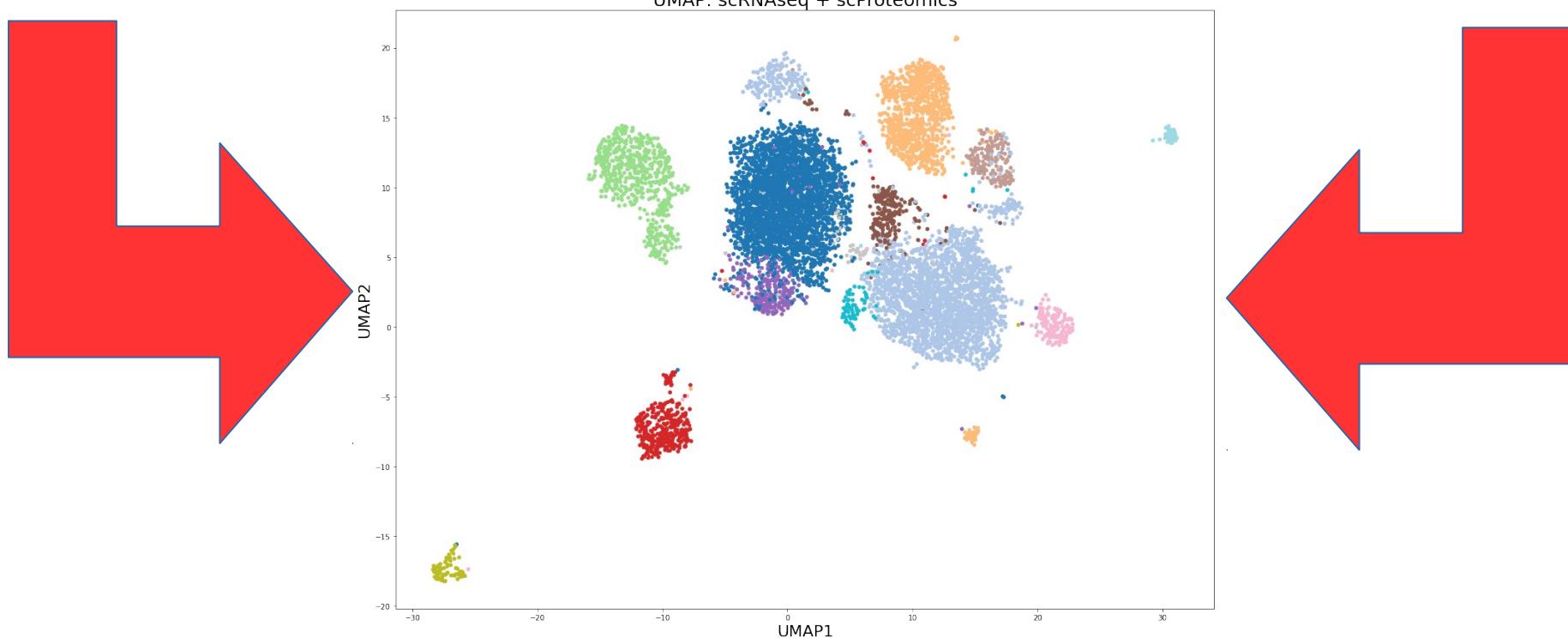
UMAP: scRNAseq



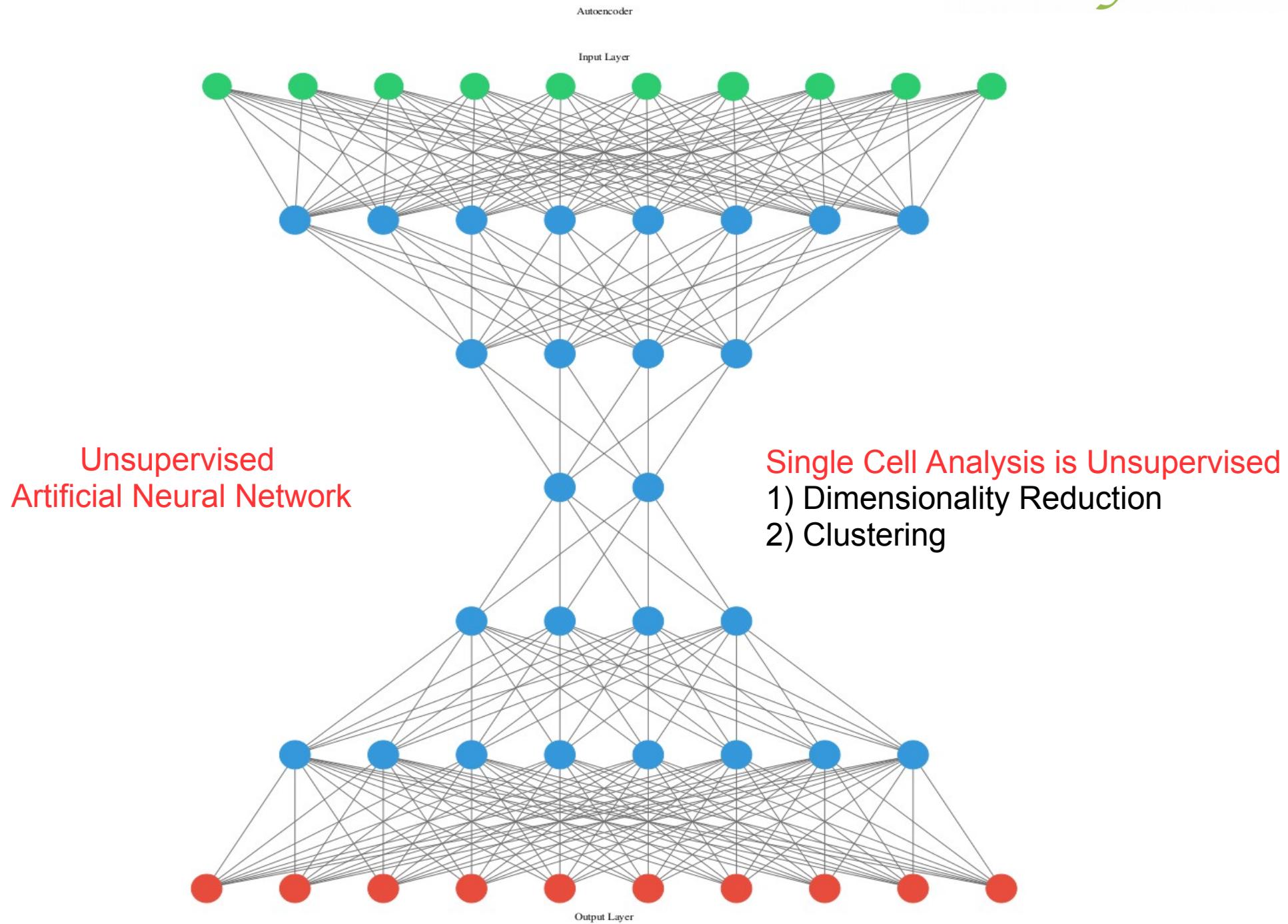
UMAP: scProteomics



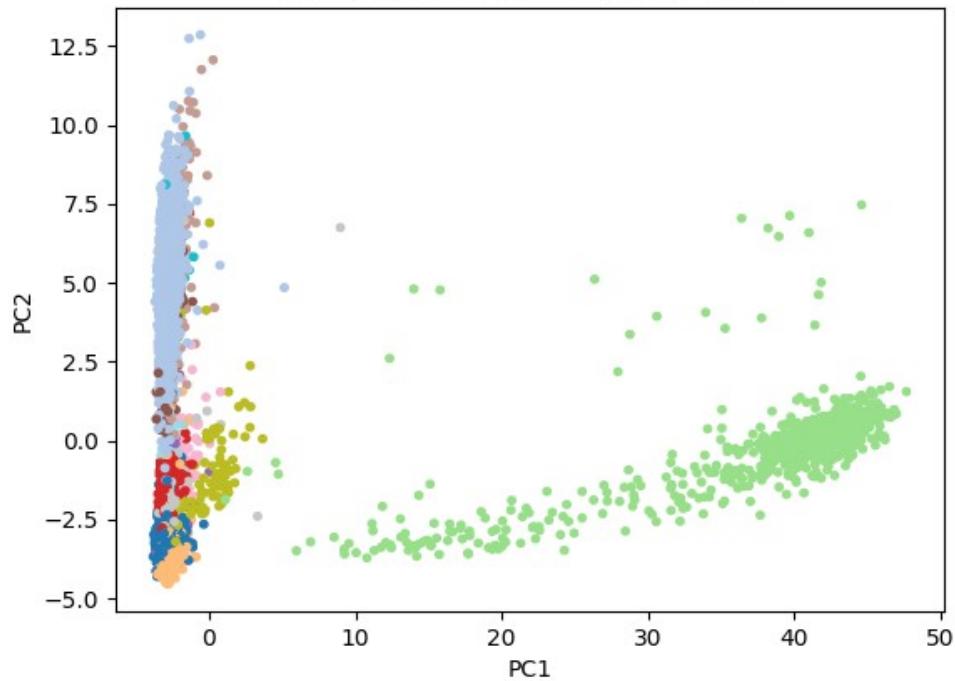
UMAP: scRNAseq + scProteomics



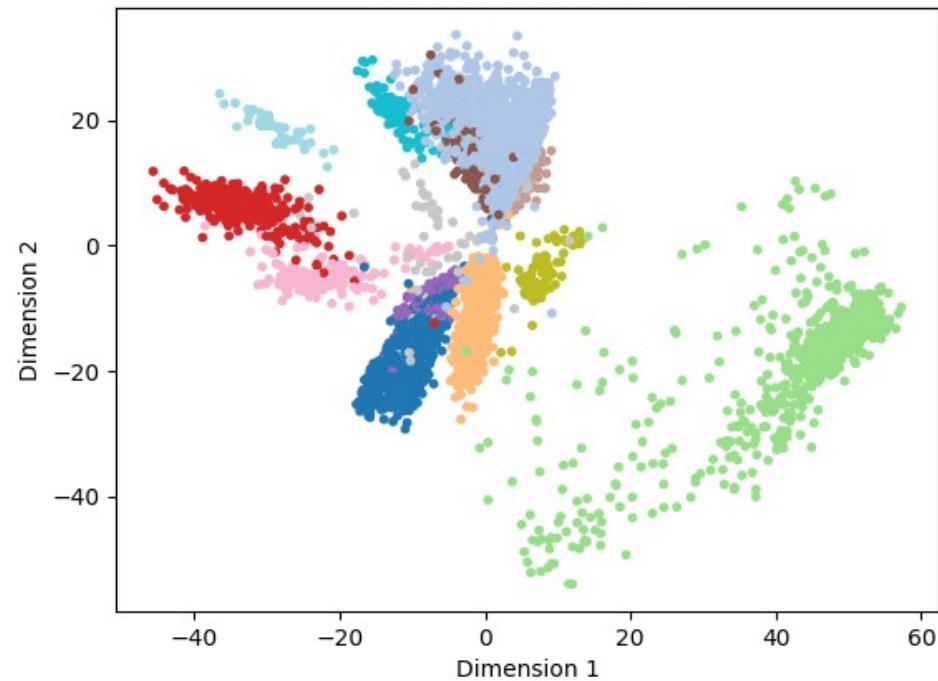
Deep Learning Integration



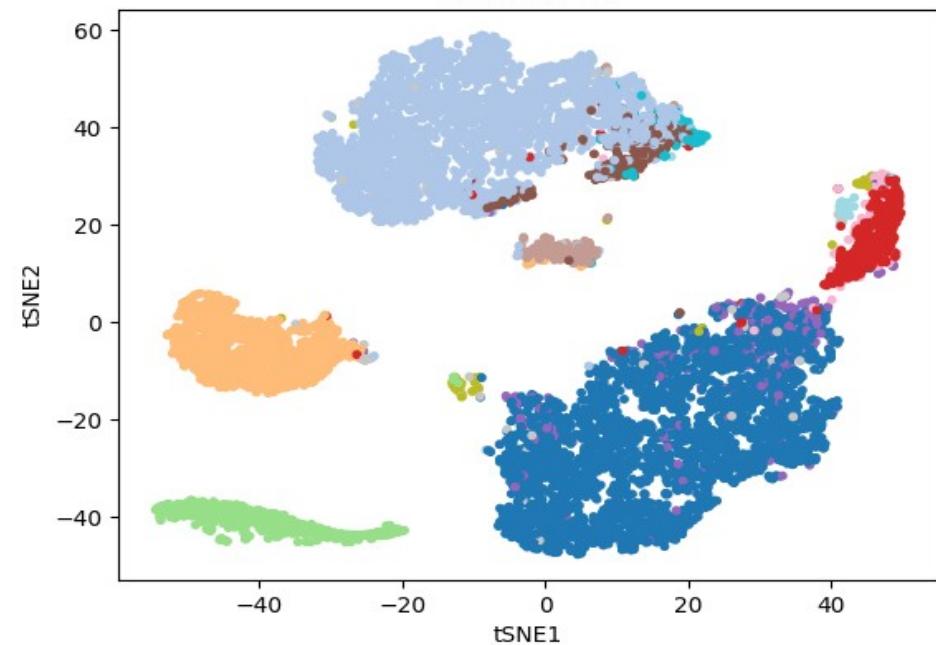
Principal Component Analysis (PCA)



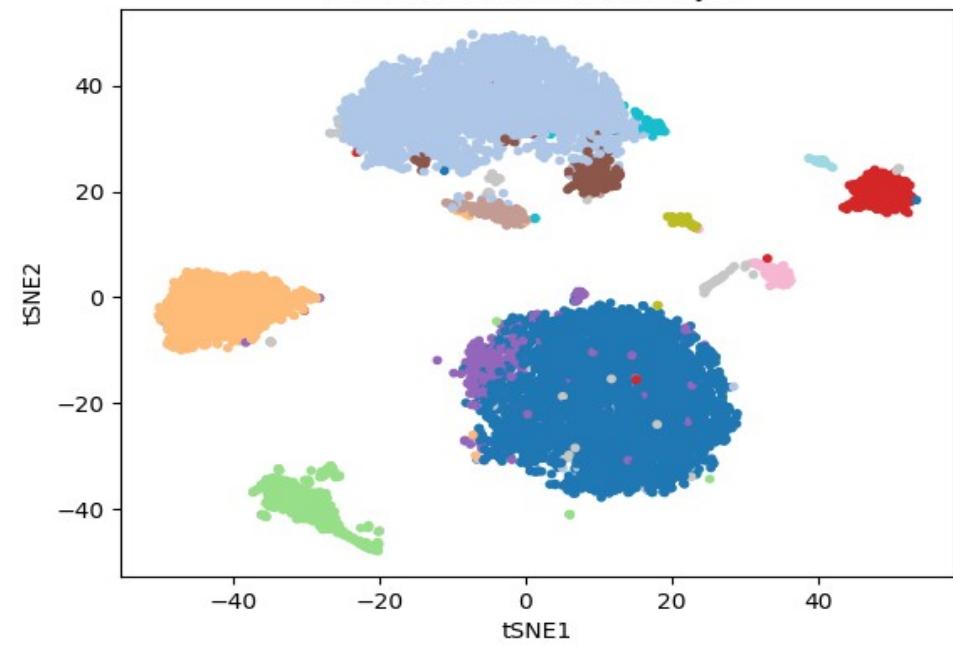
Autoencoder: 8 Layers



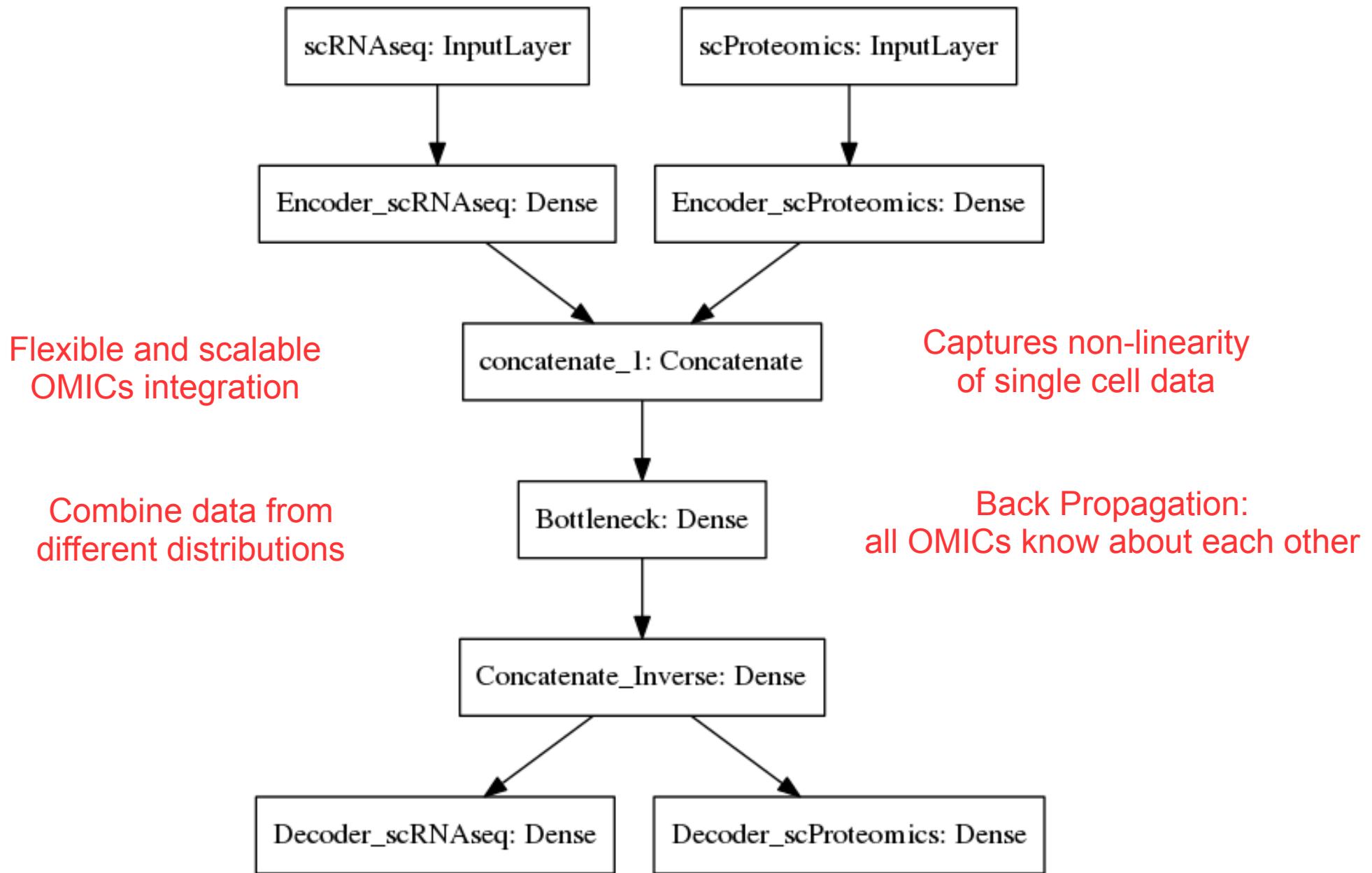
tSNE on PCA



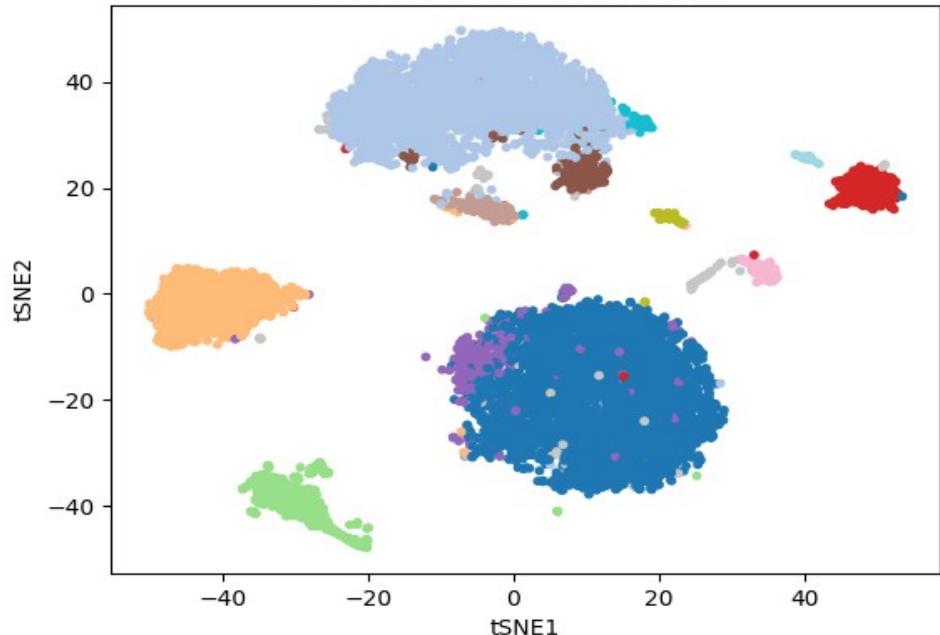
tSNE on Autoencoder: 8 Layers



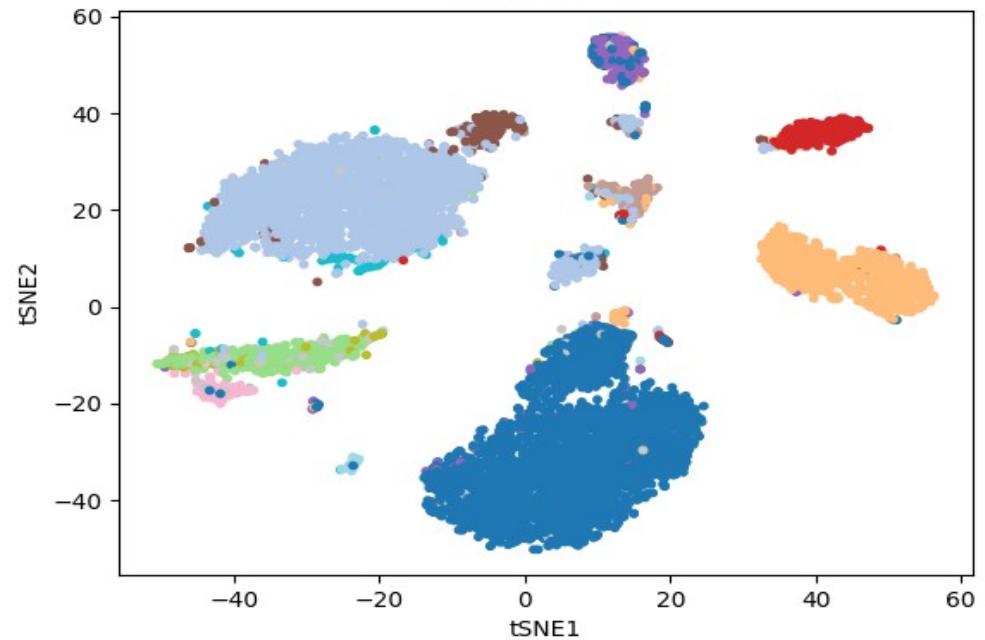
CITE-seq: Data Integration Autoencoder



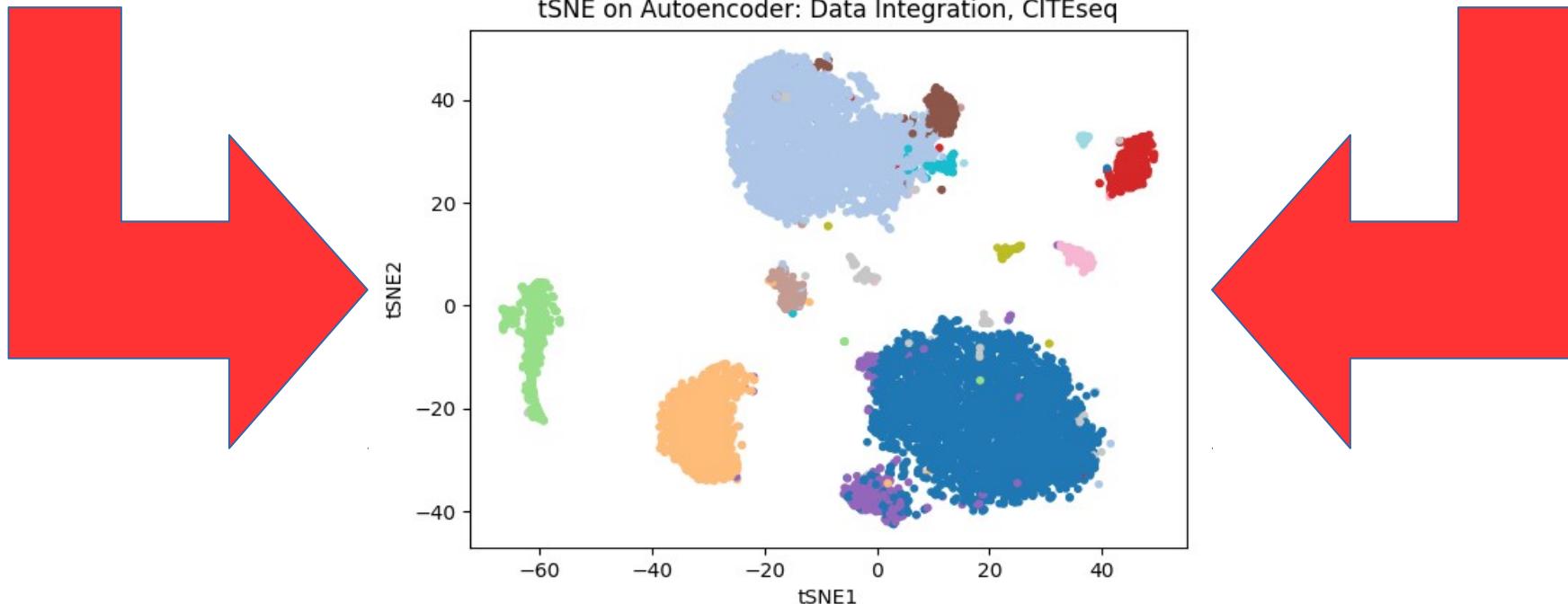
scRNAseq



scProteomics



tSNE on Autoencoder: Data Integration, CITEseq





National Bioinformatics Infrastructure Sweden (NBIS)

SciLifeLab



*Knut och Alice
Wallenbergs
Stiftelse*



LUNDS
UNIVERSITET